CHAPTER 2
THE WORKING SUN

2.1 Introduction

You have already seen that the Sun is a prodigious source of energy. The most distinctive physical property of the Sun is that it has a steady luminosity of about 3.84×10^{26} W, which is emitted mainly in the form of visible and infrared radiation from the photosphere. The aim of this chapter is to investigate some of the physical mechanisms that are associated with the Sun's energetic processes. In addition to the steady photospheric emission, the Sun is also a source of energetic outbursts; solar flares that may emit about 10^{25} J, predominantly in X-rays, over a period of a few minutes. Less obviously, there are outflows of material from the Sun, dramatic eruption events and a steady solar wind, and these too involve large releases of energy.

The first question will be the source of the solar luminosity; this will require us to consider the extreme environment that lies at the centre of our nearest star, and to review the techniques that astronomers and solar physicists use to probe regions that are hidden from our direct view. We will then move out to the surface layers of the Sun, and investigate phenomena that are linked to solar activity. The wealth of observational detail that is available about energetic processes on the surface of the Sun has helped scientists to unravel some of the processes that drive solar activity. We will see how various seemingly diverse phenomena that are indicative of solar activity can all be attributed to the behaviour of the Sun's magnetic field. We shall also consider how the Sun interacts with its environment through the outflow of material. The interaction between the Earth's magnetic field and out-flowing solar material is a source of both beauty and danger: it gives rise not only to the magnificent aurorae (the Northern or Southern Lights), but also to magnetic effects that can disrupt electrical distribution systems on the Earth. Finally, we investigate the boundaries of the Sun's dynamic environment and find that this takes us far beyond the most distant planets of the Solar System to the realms of interstellar space.

2.2 Inside the Sun

2.2.1 Introducing the solar interior

In Chapter 1 we examined in some detail the outer parts of the Sun and the radiation they emit. In particular, we saw that most of the energy received at the Earth from the Sun is carried by radiation from the photosphere. There are three factors that account for the photosphere's effectiveness as a source of energy. First, the atmospheric layers above the photosphere are, for the most part, transparent at the visible wavelengths which dominate photospheric emission. Second, the photosphere itself is opaque in the sense that although we can see a short distance into it, we cannot see through it. Third, the photosphere is sufficiently large and at a sufficiently high temperature, around 6000 K, to be a powerful thermal source of radiation.

In our quest to understand the nature and origin of the Sun's radiation, the task that now confronts us is that of explaining why the photosphere has such a temperature and how that temperature is maintained despite the prodigious rate at which solar

energy is radiated into space. You may already be familiar with the broad answer to these questions: the photosphere is heated from below by energy coming from deeper and hotter regions of the **solar interior**. This state of affairs persists because the **core** of the Sun – roughly the central 2% or 3% of the Sun's volume – is a steady and long-lived energy source powered by nuclear processes. This section is devoted to expanding these answers and giving you some feel for the nature of the solar interior.

2.2.2 The internal structure of the Sun

Discussions of the solar interior are bound to be largely theoretical because the whole region is hidden from view by the photosphere. Most of what we know about the interior is based on a number of theoretical **solar models** that have gained wide acceptance amongst solar scientists. These solar models differ from one another only in rather small matters of detail: they all agree about general principles and all give very similar results. Solar models have become increasingly refined as techniques to probe the solar interior have developed (Section 2.2.6). Indeed, by the 1990s, the reliability of solar models was such that many scientists claimed that certain anomalous results could only be explained if an aspect of our understanding of fundamental physics was incorrect – a viewpoint that we now know to be vindicated.

Each solar model is based on a few fundamental physical principles, some plausible assumptions about the interior, and some observed properties that are termed 'boundary conditions'. The *physical principles* include the requirement that the rate at which the Sun radiates energy should equal the rate at which nuclear processes produce energy in the interior, and the need for the solar material at a given depth to be able to support the weight of the matter that sits on top of it. The *plausible assumptions* relate to many issues, including the importance of internal magnetic fields and the rate at which the internal layers of the Sun rotate about the Sun's axis. The *boundary conditions* specify certain observed properties of the Sun, such as the radius ($R_\odot$), total mass ($M_\odot$), luminosity ($L_\odot$) and chemical composition.

A solar model that is constructed in this way provides numerical values for the temperature, pressure and density at any given distance from the centre of the Sun. This latter quantity is usually expressed as a fraction of the Sun's radius and is often denoted $R/R_\odot$, where $R_\odot$ is the radius of the photosphere. It is called the **fractional radius**. Representative results from a particular solar model are displayed graphically in Figure 2.1. More will be said about the route that leads to such results in the next section.

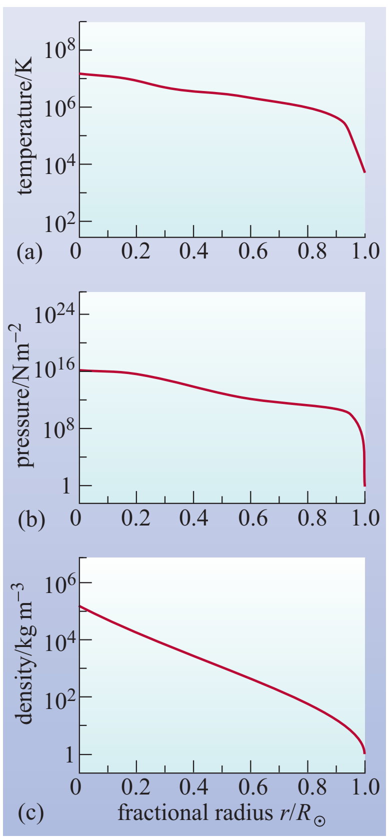


Figure 2.1 Variation with fractional solar radius of (a) temperature, (b) pressure and (c) density in the solar interior. Note that the vertical axis is logarithmic in each case.

The rise of temperature, pressure and density with increasing depth indicated by Figure 2.1 is not surprising. The pressure should be expected to increase with depth, owing to the growing weight of overlying material, and the temperature to rise, because of the increasing proximity to the central energy source. Nevertheless, the details, which emerge from a lengthy computer-based calculation, are surprising. Temperature, pressure and density all change very rapidly near the photosphere, but it's not until a fractional radius of about 0.5 that the density is equal to that of water on Earth (1.0×10^3 kg m^{-3}). Even at the centre of the Sun, where the temperature is 15.6×10^6 K, the density is predicted to be only fourteen times that of lead, though the pressure is more than 10^{10} times that of the Earth's atmosphere at sea-level.

QUESTION 2.1

Assuming that the core of the Sun occupies a fractional radius of 0.3, calculate the fraction of Sun's volume that is occupied by the core. Express your answer as a percentage.

2.2.3 The internal composition of the Sun

When dealing with the composition of the solar interior, it is traditional to divide the constituents into just three categories: hydrogen, helium and everything else. In this context, all those materials that fall into the 'everything else' category will be referred to as **heavy elements**, irrespective of precisely how heavy they are compared with hydrogen and helium. Thus elements such as boron, oxygen and carbon will be called heavy elements, along with iron, copper and gold. Accepting this convention, at least for the moment, the composition of a sample of material at any depth within the Sun can be defined by assigning numerical values to three simple parameters:

The **hydrogen mass fraction** (X):

$$X = \frac{\text{mass of hydrogen in sample}}{\text{mass of sample}}$$

The **helium mass fraction** (Y):

$$Y = \frac{\text{mass of helium in sample}}{\text{mass of sample}}$$

The **metallicity** (Z):

$$Z = \frac{\text{mass of heavy elements in sample}}{\text{mass of sample}}$$

The term metallicity arises from the fact that many astronomers use the term 'metals' to refer to all heavy elements, regardless of whether they exhibit the properties normally associated with metals. Since this usage of the term 'metals' is potentially confusing, in this book we refer instead to 'heavy elements'.

■ What can you say about the value of $X + Y + Z$ at any depth within the Sun?

❑ $X + Y + Z = 1$ at any depth. This follows from the fact that we defined Z to be 'everything else' apart from hydrogen (X) and helium (Y).

As you saw earlier, spectroscopic studies of the outer parts of the Sun can provide detailed information about the relative abundance of the various chemical elements found there. However, such information is unlikely to be a reliable guide to the constitution of the solar interior. Instead, determinations of X, Y and Z in the interior, like structural determinations, are usually based on calculations that involve a solar model.

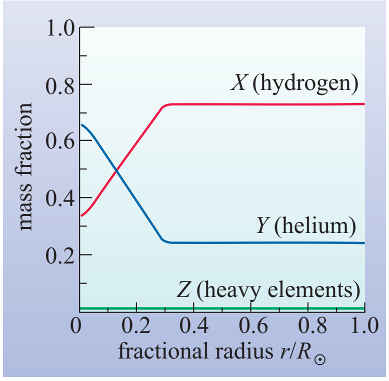


Figure 2.2 Variation of the mass fractions X, Y and Z with the fractional radius, $R/R_{\odot}$, for the Sun.

It is universally accepted that the nuclear processes that power the Sun convert hydrogen into helium. Thus, the values of the hydrogen and helium mass fractions (X and Y) change with time. Now, any solar model that provides a reasonably full account of energy production must take into account these changes, and the changes they produce in temperature, pressure and density. Consequently, a detailed solar model is capable of simulating the evolution of the Sun – including the evolution of its composition. By adjusting the initial values of X, Y and Z, until the simulated evolutionary process results in a structure that agrees with our knowledge of the present-day Sun, it is possible to obtain estimates for the current values of X, Y and Z at any depth.

The results of one such calculation of the Sun's present internal composition are shown in Figure 2.2. In this particular case, the value of Z was initially set at 0.017 and was assumed to be independent of depth and constant in time. The value of X, which was initially set at 0.735, is now expected to vary from 0.341 at the core, where a good deal of hydrogen has been converted into helium, to 0.735 at the surface, where there has been no conversion. Since $X + Y + Z = 1$, the present-day helium mass fraction, at any depth, is given by $Y = 1 - 0.017 - X$. It follows that, throughout the upper parts of the interior, where $X = 0.735$, Y retains its initial value, 0.248. This sort of result is common to many solar models and is generally taken to provide a precise determination of solar helium abundance.

To summarize, throughout most of its interior the Sun is approximately 73% hydrogen, 25% helium and 2% everything else (by mass). However, in the central 30% or so of its radius the percentage of hydrogen is increasingly depleted (and helium compensatingly increased), with about half the hydrogen initially present in the core of the Sun having been converted into helium.

Before leaving the subject of the Sun's composition, one point, mentioned earlier, deserves special emphasis. Most of the atoms in the Sun are *ionized*. This is particularly true in the hot, dense interior, where essentially all the hydrogen and helium atoms are completely ionized. Such a highly ionized gas is called a **plasma**. So, although it is quite common to see the Sun referred to as a gaseous body, a more specific description is that it is made of plasma. In this case, the plasma consists mainly of hydrogen and helium ions, together with the electrons that were liberated when those ions were produced.

QUESTION 2.2

Assuming that solar material pretty much stays in one place during the life of the Sun (that is, there is no significant internal mixing of solar constituents), what can you deduce so far about the location of the nuclear processes that convert hydrogen into helium?

Indeed, the location of the nuclear processes in a central core is a *direct* and continuing consequence of the increase of temperature with depth in the models: the rates of nuclear reactions rise very rapidly as the temperature increases.

2.2.4 The energy source of the Sun

This section addresses two questions:

- What are the main energy-releasing nuclear processes that take place in the core of the Sun?
- At what rate do those nuclear processes occur?

The answers to these questions provide the ultimate solution to our search for the true origin of the Sun's electromagnetic radiation. But, as you will see, they also raise other questions.

What are the energy-releasing nuclear processes?

The basic energy-releasing process taking place in the Sun is *nuclear fusion*. This is the process in which nuclei of relatively low mass are fused together to form nuclei of somewhat greater mass. The fusion is brought about by a sequence of nuclear reactions in which colliding nuclei combine and fragment, to produce new nuclei together with other particles. No energy is actually created in these reactions: it is simply that energy is liberated from the *reactants* and is redistributed amongst the *products* in such a way that some of it replaces the energy radiated by the Sun. This replacement of lost energy maintains the high temperature of the core, thus sustaining the nuclear reaction rates.

A nuclide is a nucleus with a particular atomic number, Z, and mass number, A.

A full account of the nuclear processes taking place in the Sun would be very complicated indeed. Our discussion will be limited to the one process that is thought to be responsible for the bulk of the Sun's radiant energy – the so-called **ppI chain** (note that the 'I' in ppI is the roman numeral 'one'). The details of this process were first described by Hans Bethe and Charles Critchfield in the 1940s. The name 'pp' indicates that it is the first of several different chains of reactions that start with colliding protons – p is the symbol for the proton, as is ${}^{1}_{1}\mathrm{H}$, the proton being the nucleus of the common nuclide of hydrogen. The nuclei involved in the ppI chain are the hydrogen nuclides ${}^{1}_{1}\mathrm{H}$ and ${}^{2}_{1}\mathrm{H}$ (which is called deuterium) and the helium nuclides ${}^{3}_{2}\mathrm{He}$ and ${}^{4}_{2}\mathrm{He}$. The other particles that are involved are of three types:

γ-rays (denoted by γ) These are just energetic photons of electromagnetic radiation, a concept that should already be familiar from the discussions in Boxes 1.1 and 1.3.

Positrons (denoted by e^+) These are particles similar in many ways to electrons (denoted by e^-); they have the same mass for instance. However, some of their properties are radically different. Of particular importance is the fact that positrons have positive charge (hence their name) whereas electrons have negative charge. Positrons are sometimes called *anti-electrons*.

Neutrinos (denoted by ν, pronounced 'new') These are electrically neutral particles (hence the name) which have a very low mass. Neutrinos travel at essentially the speed of light and interact with other particles so weakly that they are able to travel through ordinary matter with great ease. Day and night, enormous numbers of neutrinos stream through the Earth with hardly any impediment. While you are reading this sentence, more than a million million neutrinos will pass through your own head. There are three types of neutrino, and it is the type called an 'electron neutrino' (ν_e) that is created in the ppI reaction.

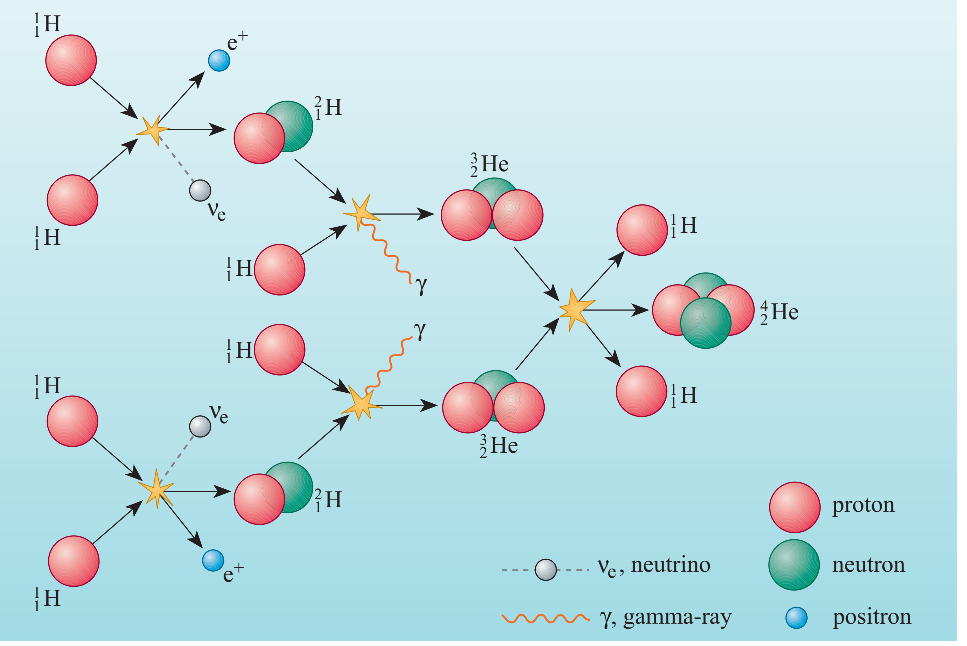


Figure 2.3 The ppI chain of nuclear reactions that is predominantly responsible for the conversion of hydrogen into helium in the Sun. Note that six hydrogen nuclei are required to initiate the chain but two are released again at the end.

The various steps in the ppI chain are shown in Figure 2.3. Pruned of its details, the overall effect of the chain is the following:

$$4\,^1_1\text{H} \rightarrow \,^4_2\text{He} + 2\text{e}^+ + 2\nu_\text{e} + 2\gamma$$

Thus, four protons are consumed and a helium nucleus containing two protons and two neutrons is produced along with two positrons, two neutrinos and two γ-rays. Each occurrence of the ppI chain is accompanied by a number of supplementary reactions that result, amongst other things, in the annihilation of the two positrons (along with two of the Sun's many electrons – this reaction does, of course, conserve charge), and the release of yet more γ-rays. So, apart from the production of helium nuclei and neutrinos, the final outcome of the ppI chain is the release of γ-rays. It is these γ-rays that are the ultimate source of much of the Sun's electromagnetic radiation.

At what rate do the nuclear processes occur?

We have now identified the process that is mainly responsible for the Sun's radiant energy. How common is this process? In order to get a rough idea of the answer to this question all we have to do is to divide the Sun's luminosity (the energy radiated per second) by the radiant energy liberated by each occurrence of the ppI chain. Making that estimate is our next goal. To start with, we need to know the energy released per occurrence.

■ Why does the above procedure provide only a rough idea of the rate at which the ppI chain occurs?

❏ Because we are neglecting all the other nuclear reaction chains, apart from the ppI chain, that also contribute to the generation of energy in the Sun.

HANS BETHE (1906–2005)

Hans Bethe (Figure 2.4) was born in Strasbourg (which was in Germany at the time, and is now in France) and educated at the Universities of Frankfurt and Munich. After the Nazi party came to power in 1933, Bethe lost his post at Tübingen on account of his mother being Jewish. He worked briefly in Britain before moving on to Cornell University in the United States in 1935.

During the 1930s it was suspected by most physicists that the key to the source of energy of the Sun must be some type of nuclear reaction. Bethe originally suggested that hydrogen might undergo fusion in a process in which carbon acts as a catalyst – a process that is now called the CNO cycle and is recognized to be important in stars of higher mass than the Sun (Chapter 6). The details of the proton-proton chain were discovered a few years later by Bethe in collaboration with Charles Critchfield. Bethe's contribution to understanding the way in which stars are powered was recognized by the award of the Nobel Prize for physics in 1967.

Figure 2.4 Hans Bethe. (Robert Barker/Cornell University)

All nuclear reactions are constrained by a number of regulating principles called *conservation laws*. These dictate which reactions are possible and determine how much energy the various particles may carry away from a reaction. Two quantities that are conserved in nuclear reactions are the electric charge (which is usually expressed as a multiple of the charge on a proton e) and the baryon number. The term **baryon** refers to a family of subatomic particles that includes the proton and the neutron, but does *not* include electrons, positrons, neutrinos or photons. The baryon number of each proton and each neutron is +1, and is zero for any particle that is not a baryon. The following example illustrates these two conservation rules in action.

EXAMPLE 2.1

Consider the initial nuclear reaction of the ppI chain as shown in Figure 2.3.

(a) Write out this reaction symbolically using the notation introduced in the discussion of the overall effect of the ppI chain.

(b) For the reaction you have written down, work out the total amount of electric charge entering and leaving the reaction, and confirm that the reaction obeys the principle of conservation of electric charge.

(c) Work out the total baryon number entering and leaving this reaction, and confirm that the principle of conservation of baryon number is obeyed.

SOLUTION

(a) The initial step of the ppI chain involves a reaction between two protons ($^1_1\mathrm{H}$) and results in the formation of the hydrogen nuclide ($^2_1\mathrm{H}$) as well as a positron and a neutrino. This can be written symbolically as:

$$2\,^1_1\mathrm{H} \rightarrow {}^2_1\mathrm{H} + \mathrm{e}^+ + \nu_\mathrm{e}$$

(b) The incoming electric charge is that of the two hydrogen nuclei. Since each $^{1}_{1}\mathrm{H}$ nucleus is simply a proton, its charge is e. Thus the total charge of the two incoming nuclei is $2e$. The outgoing hydrogen nucleus, $^{2}_{1}\mathrm{H}$, carries charge e (as indicated by the subscript, the atomic number) and the positron carries charge e. Thus the total outgoing charge is $2e$ and hence electric charge is conserved.

(c) The baryon number entering or leaving is equal to the total number of protons and neutrons entering or leaving, and this total is given by the superscript, the mass number. The baryon number of the incoming nuclei is +2 (each has a baryon number of +1). The total baryon numbers of the products are +2 (for $^{2}_{1}\mathrm{H}$), 0 (for the positron) and 0 (for the neutrino), giving a total baryon number of +2. Thus the baryon number is conserved in this reaction.

QUESTION 2.3

Example 2.1 considered the first nuclear reaction in the ppI chain. For the remaining two reactions:

(a) Write out each of the reactions symbolically.

(b) Show that these reactions obey the principle of conservation of electric charge.

(c) Show that these reactions obey the principle of conservation of baryon number.

Just as nuclear reactions conserve electric charge and baryon number, so they must also obey another principle, that of *conservation of energy*. The total energy emerging from a reaction must be equal to the total energy that entered the reaction. In applying this principle to nuclear reactions, it is important to include any **kinetic energy** (i.e. energy by virtue of motion) that the particles may have, but it is also important to remember Einstein's famous discovery that even particles at rest have a certain amount of energy, called the **rest energy**. The rest energy of a particle of mass m is given by

$$E = mc^2 \tag{2.1}$$

By using this formula, together with the principle of conservation of energy, it is possible to estimate the amount of radiant energy ultimately released (as γ-rays) by each occurrence of the ppI chain. In order to make the estimate we shall assume that, once all the supplementary reactions are taken into account, there is no significant change in the kinetic energy of the particles present before and after each occurrence of the ppI chain. We shall also assume that the energy associated with the two neutrinos is negligible (even though it's actually about 2% of the total). With these assumptions, we have to consider only the rest energies involved in the ppI chain – though we must remember to add the contribution arising from the annihilation of the two positrons with two of the Sun's electrons to produce yet more radiant energy. Thus,

$$\begin{array}{c}\text{radiant energy eventually}\\ \text{resulting from each}\\ \text{occurrence of the ppI chain}\end{array} = [(4 \times \text{rest energy of } {}^{1}_{1}\mathrm{H}) - (\text{rest energy of } {}^{4}_{2}\mathrm{He} + 2\mathrm{e}^{+}) + (\text{rest energy of } 2\mathrm{e}^{+} + 2\mathrm{e}^{-})]$$

Having got this far, there is no reason why you shouldn't do the rest of the work for yourself, so here's your chance.

QUESTION 2.4

The mass of the helium nucleus, ${}^{4}_{2}\mathrm{He}$, is 6.645×10^{-27} kg. That of the hydrogen nucleus, ${}^{1}_{1}\mathrm{H}$, is 1.673×10^{-27} kg, and that of each electron or positron is 9.110×10^{-31} kg.

(a) Use this information to find the radiant energy eventually resulting from each occurrence of the ppI chain, under the assumptions given above.

(b) Use your answer to part (a), together with the value of the solar luminosity of $3.84 \times 10^{26}\,\mathrm{J\,s^{-1}}$, to estimate the rate at which the ppI chain occurs.

(c) Just for fun, use your answer to part (b) to estimate the mass of hydrogen consumed per year by the ppI chain.

The Sun is such a massive body – 1.99×10^{30} kg – that, despite the consumption of an enormous amount of hydrogen every year, it has been able to shine fairly steadily for about 4.5×10^{9} years and will probably continue to do so for another 4 or 5×10^{9} years.

The above discussion should have given you a clear general picture of the main processes leading to the release of energetic γ-rays in the hot, dense conditions of the Sun's core, and hence of the origin of solar radiation. However, you should also be aware that there are some major questions that still need to be answered. A fairly obvious one that arises whenever conditions in an inaccessible region are described is 'how do you know it's really like that?' This will be addressed shortly in the section entitled 'Testing theories of the solar interior'. But an even more pressing question is this: 'How does the energy liberated in the core of the Sun account for the visible light that emerges from the Sun's surface?' This issue, which relates directly to our concern about how the Sun shines, is the subject of the next section.

2.2.5 How energy reaches the surface

The core of the Sun is a hot dense plasma – more than ten times denser than metals in our everyday world. Photons are not likely to travel very far through this material before they encounter an electron or an ion. Therefore the γ-rays produced by the ppI chain and other solar nuclear reactions are quite unable to travel directly to the Sun's surface and onward into space. Instead, there must be some mechanism, or set of mechanisms, whereby energy is *transported* from the core of the Sun to its surface. There are three fundamentally different methods of transferring energy from place to place that might, in principle, be involved – *radiation, conduction* and *convection*. Before explaining which are important in the Sun, we shall examine the basic principles of all three as a piece of background science in Box 2.1.

BOX 2.1 METHODS OF ENERGY TRANSFER: RADIATION, CONDUCTION AND CONVECTION

Radiation

In the case of **radiation**, the energy is carried from place to place by waves, rays or streams of particles that are emitted and absorbed. We usually think of *radiative energy transport* in terms of electromagnetic radiation, as in the case of the infrared radiation from an electric fire, or the light from the Sun. However, radiative energy transport is not really quite so limited. In some circumstances, for instance, streams of neutrinos can transfer substantial amounts of energy. A characteristic of radiation is that it may operate across a vacuum, though it does not necessarily have to do so.

Conduction

In contrast to radiation, **conduction** can transfer energy only between places that are physically linked by a material medium. In conductive energy transport the basic mechanism is that moving particles collide and redistribute their energy, and a common characteristic is the existence of a temperature gradient between the source of the energy and its destination. A common example of conduction is the heating of a metal spoon used to stir hot soup. The soup heats the end of the spoon with which it is in contact, causing the atoms in the metal to vibrate rapidly. The energy of these vibrations is gradually passed (mainly by electrons) to more slowly moving neighbouring atoms in cooler parts of the spoon. In this way the cooler parts are heated and some of the energy initially contained in the soup is eventually transferred to the handle of the spoon.

Convection

Convection, like conduction, requires a medium. Moreover, in the case of convection, the medium must be a fluid (a gas, a liquid or a plasma) and it must be in a gravitational field. *Convective energy transport* can also be exemplified by a common culinary experience, that of heating a saucepan of water. When a saucepan is placed on a gas ring or an electric hob the water at the bottom of the pan is heated and expands. This reduces the density of the water at the bottom of the pan, so, under the influence of the Earth's gravitational field, it starts to rise, displacing the cooler denser water above. When the heated water reaches the surface it radiates away some of its energy and cools down while the water at the bottom of the pan is warmed by more energy coming (by conduction) from below. Thus, the cycle is able to repeat itself and a pattern of **convection currents** is established within the saucepan – as indicated in Figure 2.5. The process is easy to see if there are objects such as peas in the water.

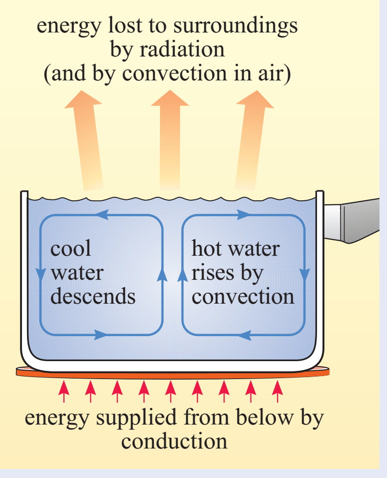


Figure 2.5 Radiation, convection and conduction in and around a heated saucepan of water.

Given the high density of the deep solar interior, you might expect that conduction would be important there, but this turns out not to be the case. The main mechanism of energy transport throughout the central 70% or so of the Sun's radius is actually radiation involving photons. Of course, the radiative energy transfer that takes place is not the sort with which we are familiar. Because of the conditions there, the photons have to make their way through the solar material via a lengthy sequence of encounters with other particles, in which they are either scattered, or absorbed and re-emitted. The photons emerging from each of these encounters have an almost equal chance of travelling in any direction, so their progress up to the photosphere

takes the form of a **random walk**, as indicated schematically in Figure 2.6. A typical step in the random walk is only a few centimetres long (even less in the core), so the outward spread of energy from the core is a gradual 'diffusive' process.

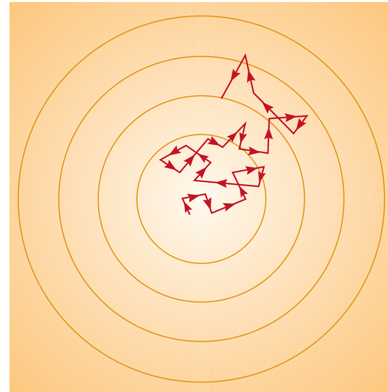


Figure 2.6 Radiant energy emerging from the solar core initially makes its way outwards via a series of radiative processes along a random walk. The graduated tone represents increasing distance from the Sun's centre. For clarity, this schematic diagram shows the path of one photon after every interaction. In reality, more than one photon may be produced after each interaction.

■ In terms of the electromagnetic spectrum, what is the difference between photons that are produced in the core of the Sun and the majority of photons that are produced at the photosphere?

❑ The photons that are generated in the core of the Sun are typically in the γ-ray part of the electromagnetic spectrum, whereas the majority of the photons emitted from the surface correspond to the visible part of the spectrum.

Since the rate of energy generation in the core is the same as the rate at which energy is lost from the photosphere, the degradation of the average photon energy between the core and the surface implies that the number of photons must increase.

There is a two stage process by which the average energy of photons is degraded from the very high energies that result from the ppI chain to the relatively low energies associated with visible light. The first stage is that the γ-rays that are generated by nuclear reactions undergo multiple scatterings with the electrons and ions in the core. In general, each scattering will redistribute energy between a photon and the electron or ion that it scatters off. This has the effect of diminishing the initially very high photon energy, and transferring energy to the particles of the plasma.

While it would be impractical to predict what would happen to an individual photon after every single scattering event, it is possible to determine what the distribution of photon energies would be for many photons that all undergo multiple scattering events. In fact, you have already encountered this distribution of photon energies.

■ Photons which are generated in the core are much more likely to interact with electrons or ions within the core than to escape. What type of spectrum is likely to result from such conditions?

❑ A black-body spectrum. In Box 1.2 it was stated that a source in which photons are much more likely to interact with the material within the source than to escape is a condition for the formation of a black-body spectrum.

So the process of multiple scattering changes the distribution of photon energies from the γ-rays that are produced by the ppI chain to a black-body spectrum that is characteristic of the core temperature. This process is called **thermalization**. However there is another key aspect to this process: the electrons and ions within the plasma interact with each other to *produce* photons. Again, because of the high degree of interaction between matter and radiation, these photons also have a spectrum that corresponds to a black-body source at the core temperature. The overall effect is that the energy that was generated as a small number of high-energy photons is now in the form of a large number of lower energy photons whose energies are distributed according to a black-body curve. This, of course, is a black-body spectrum that corresponds to the core temperature of the Sun, and has a peak at much shorter wavelengths than the approximately black-body spectrum that is emitted from the photosphere.

QUESTION 2.5

The ppI reaction produces positrons, which undergo annihilation reactions with electrons. Each annihilation event yields two γ-rays, each of which has an energy of 0.51 MeV (1 MeV = 10^6 eV).

(a) Calculate the wavelength of the γ-ray photons produced by positron–electron annihilation.

(b) Calculate the peak wavelength of the Planck curve that corresponds to the core temperature of the Sun. In what part of the electromagnetic spectrum does this peak lie?

It should also be noted that not all of the energy released by the pp chains is in the form of γ-rays. A substantial fraction of the energy that is liberated in each reaction is in the form of the kinetic energy of the nuclei or particles that are formed. Any nuclei that are formed soon scatter off protons and electrons in the dense plasma in the core. The kinetic energy of one particle is soon redistributed to many particles moving at random in the plasma, and the energy is converted into the thermal energy of the plasma. This is also a process of thermalization, but the term now applies to nuclei or particles rather than photons. The one product of the pp chains that does not share energy with the plasma in the core is, of course, the neutrino. Neutrinos simply stream out of the Sun without interaction while carrying with them a small fraction of the energy that is liberated by the pp chains.

The second stage of the degrading of the average energy of photons arises from the gradual nature of the outward transport of energy from the core to the surface of the Sun. As we have seen, basic physical principles, the principles that provide the foundation for a solar model, lead to the conclusion that the interior temperature of the Sun decreases as the distance from the centre increases. This decrease in temperature is reflected in the distribution of speeds found amongst electrons or ions of a given type at various distances from the Sun's centre. On the whole, the lower the temperature in any given region, the lower the average speed of particles (of a given type) in that region. As photons diffuse outwards they are scattered and absorbed by interactions with electrons and ions, with the result that the energy of the average photon is gradually reduced. No energy is lost in this process; it is simply redistributed and shared amongst increasing numbers of photons. The sharing and redistribution of energy is highly effective and the local electrons and ions may be regarded as the immediate source of the radiation. This state of affairs is described by saying that the radiation is in **local thermodynamic equilibrium** with the material through which it passes. This is in fact a more formal way of stating the condition about the interaction between radiation and matter that is necessary for the formation of a black-body spectrum. Thus in a region that has a certain temperature, the electromagnetic radiation has a black-body spectrum that is characteristic of that temperature. This applies right up to the photosphere.

Detailed calculations show that, under the conditions of temperature and density found in the solar interior, convection is the main mechanism responsible for energy transport throughout the outer 30% or so of the Sun's radius. In the solar context, it is very strongly suspected that the convection currents are divided into a number of **convection cells**. The uppermost cells are quite small, typically 1000 km across, but quite deep. They account for the ever-changing solar granulation seen in the

photosphere. (This was described in Section 1.2.3.) A deeper layer of larger cells, typically 30 000 km across, is thought to account for a similar but less obvious phenomenon called **supergranulation**. Supergranules are not seen as light and dark patches on the photosphere, but their presence can be deduced from detailed observations of the movement of photospheric material or from various magnetic field measurements. It has also been suggested that there might be an underlying layer of giant convection cells, but there is no strong evidence in favour of this proposal.

Figure 2.7 shows a cross-section of a sector of the Sun and is designed to emphasize energy transport. As you can see, beyond the core (where the nuclear reactions take place), it is conventional to call the region in which energy is still mainly transported by radiation the **radiative zone** and the outer region in which convection dominates the **convective zone**. The top of the convective zone roughly corresponds to the photosphere, where radiation once again becomes the dominant mechanism for energy transport. Figure 2.8 shows the location and relative sizes of these regions in a cut-away view of the Sun.

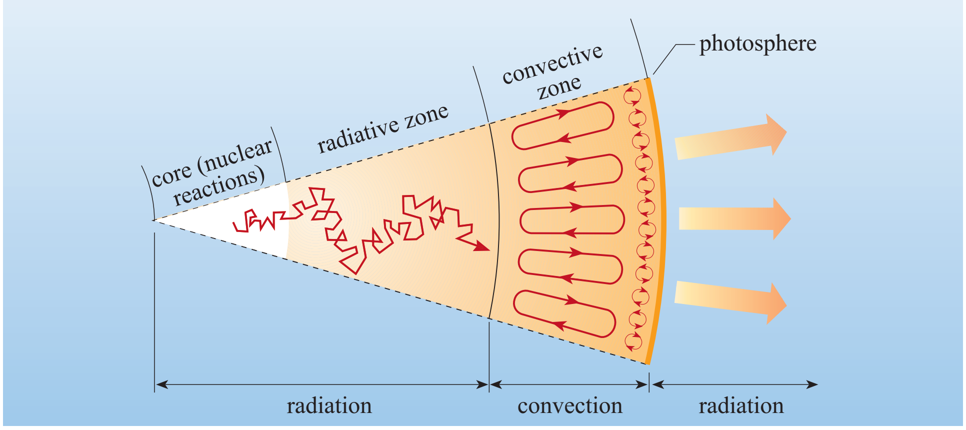


Figure 2.7 Energy transport in the Sun.

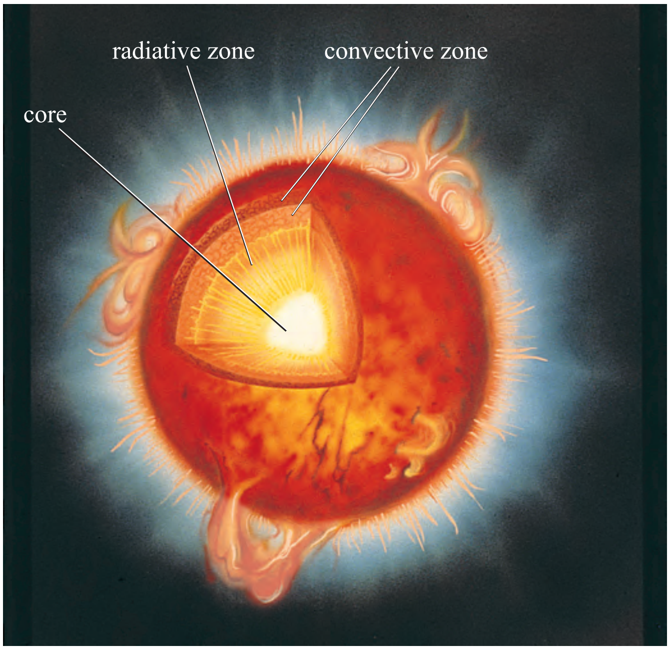


Figure 2.8 A cut-way view of the solar interior showing the location and relative sizes of the core, the radiative zone and the convective zone.

Light takes about 8.3 minutes to travel from the photosphere to the Earth. If the Sun were entirely transparent it would only take radiation about 2.3 seconds to make the journey to the photosphere from the core. However, the Sun is not transparent and the journey actually takes a lot longer. It has been estimated that energy being released in the core right now will take about two hundred thousand years to reach the surface. Fortunately, as you will learn in the next section, there is evidence that energy is still being released in the Sun's core.

2.2.6 Testing theories of the solar interior

Are our ideas about the Sun's interior right? Of course, theories are constantly under review; theorists are engaged in an unending search for inconsistencies or unseen implications within existing ideas, while at the same time looking for possible applications of new ideas. There is also a constant interplay of theory and experiment as new laboratory measurements of quantities such as nuclear reaction rates permit improved calculational precision. But, ultimately, theories must be tested against observation. It is just such observational tests that are the subject of this section.

In the case of the solar interior, current ideas can be subjected to a number of observational tests, though some are less conclusive than others. For example, some success has been achieved by those who attempt to explain the observed properties of Sun-like stars on the basis of modified solar models, but observations of other stars inevitably lack the detail and precision that are easily obtained in solar observations. Important information about the solar interior has been obtained from observations of the Sun's overall shape and surface composition, but even these data fail to provide much insight into the properties of the deep interior. Despite this catalogue of woes, recent years have witnessed two major developments that have already had an enormous influence on the study of the Sun. These two developments will now be described.

Observations of solar oscillations

It has been known since the early 1960s that the Sun oscillates – its surface moves up and down. The maximum speed of the surface during these oscillations is about 500 m s^{-1}.The observed motions are not, at first sight, particularly orderly – localized regions of the photosphere rise and fall, somewhat irregularly, through distances of many kilometres in characteristic time periods of five minutes or so. Despite the lack of apparent coherence it was established in the mid-1970s that the observed movements partly result from the combined effect of many simple **global oscillations** that individually are very orderly indeed. A few of these global oscillations are illustrated in Figure 2.9. Each involves a coherent movement of the entire solar surface, and each has its own characteristic time period. The contribution of an individual oscillation to the overall motion of the surface is tiny – typically less than 0.2 m s^{-1}. Some of the low-frequency global oscillations actually penetrate deep into the solar interior, as indicated in Figure 2.10. In view of this, it is not surprising that conditions in the Sun's interior influence the relative significance of the various global oscillations and thus the detailed surface movements they jointly produce.

Thanks to the existence of these deep-rooted global oscillations it is possible to learn about the Sun's interior by observing its surface. Because this is similar to the way that terrestrial seismologists learn about the Earth's interior by studying the

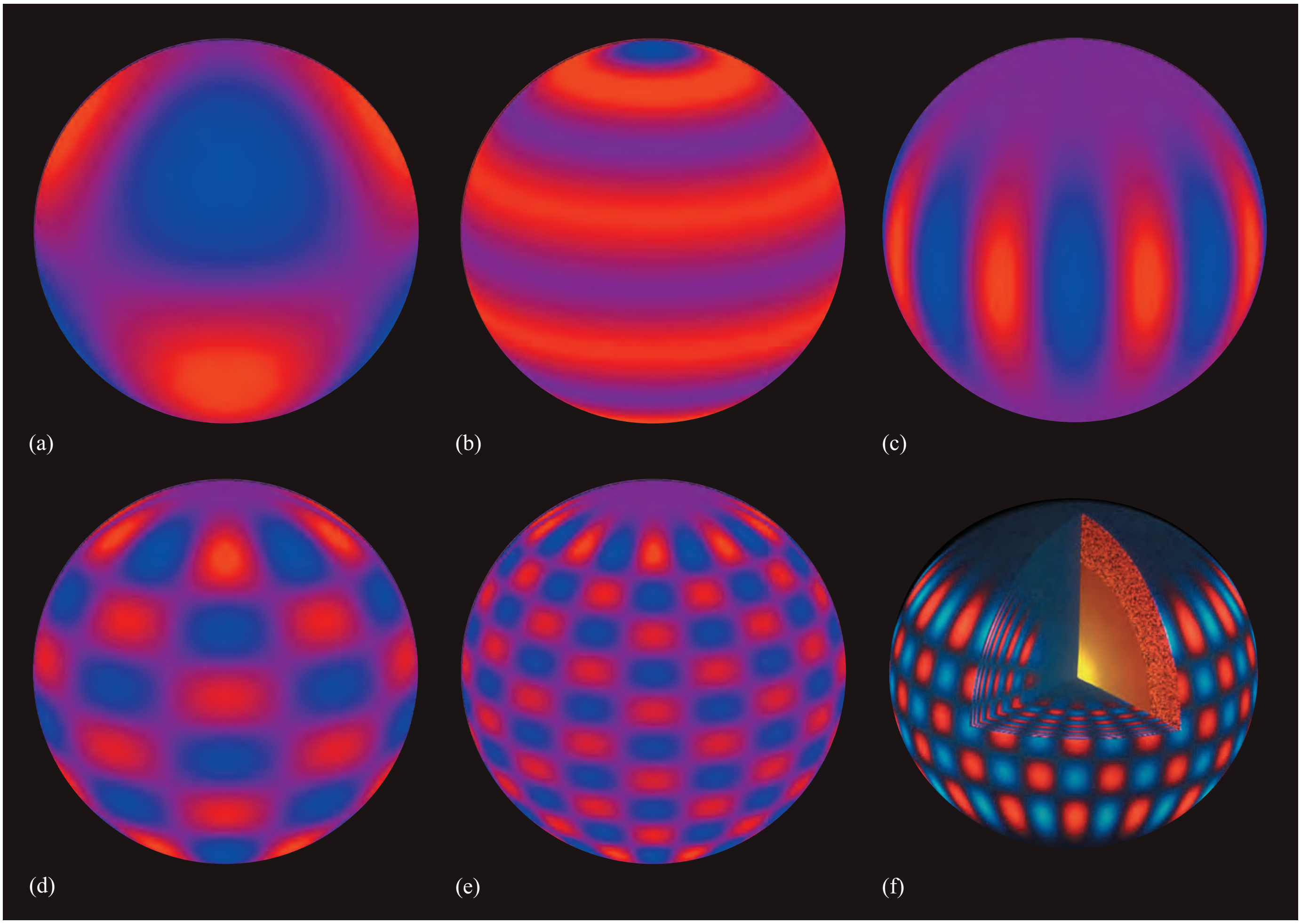


Figure 2.9 Global oscillations of the Sun. The red areas represent zones of temporary expansion, the blue areas show zones of temporary contraction. Figures (a) to (e) show the motion of the surface of the Sun for different modes of oscillation. (f) shows the surface and a schematic cut-away illustrating how one particular mode propagates within the Sun. ((f) National Solar Observatory)

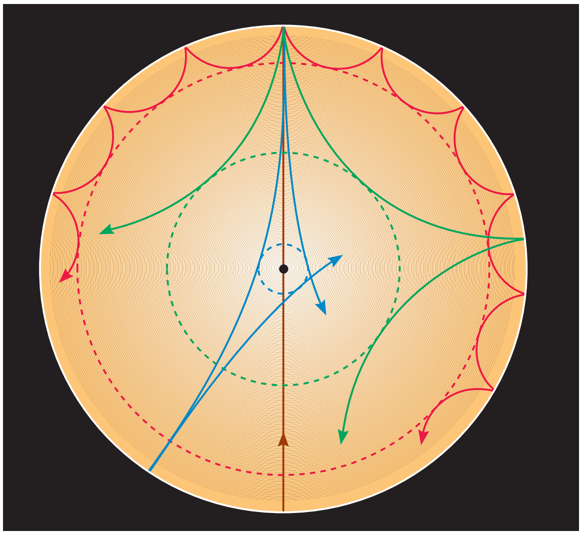


Figure 2.10 Different global oscillations will travel different distances into the Sun. This diagram shows how four modes of oscillation propagate within the body of the Sun. Some modes, such as that shown in red here, can only reach shallow depths within the Sun before being refracted back towards the surface. Other modes are capable of reaching deep within the Sun, and some, such as the modes shown here in blue and brown reach the core. (Adapted from Lang, 2001, from an original figure by J. Christensen-Dalsgaard and P. H. Scherrer.)

vibrations initiated by earthquakes, the subject has come to be called **helioseismology**. It is also possible to study, in a rather limited way, global oscillations in stars apart from the Sun, and this is referred to as **asteroseismology**. In the case of the Sun and stars, there are a variety of sources that act as the initial driver of oscillations, the most prominent being the constant churning motion of the convective zone.

Significant advances in helioseismology were made during the 1990s as a result of specialized programmes in which the Sun is observed continuously for weeks or even months at a time. Such observations are necessary if the properties of solar oscillations are to be measured well enough to be used for helioseismology.

- Why is a single ground-based telescope likely to be unsuitable for making such long continuous observations?
- Observations from a single ground-based telescope will be interrupted on a daily basis as the Sun sets for the night. (Of course, this condition would not apply if such measurements were made from polar regions, where the Sun might not set for weeks on end. Observing the Sun from the South Pole has been attempted but is hampered by practical difficulties.)

In order to make continuous long-term measurements of the Sun, two different approaches have been used. A simple, yet expensive, strategy is to place the telescope in space at a location that allows an uninterrupted view of the Sun. This was the approach taken for the multinational SOHO (SOlar and Heliospheric Observatory) mission that was launched in 1995. A different approach is to construct a network of telescopes at different longitudes around the Earth, such that by the time the Sun sets on one telescope, it has risen, and is being monitored from, another telescope in the network. Several such networks exist, and provide very long-term solar monitoring programmes.

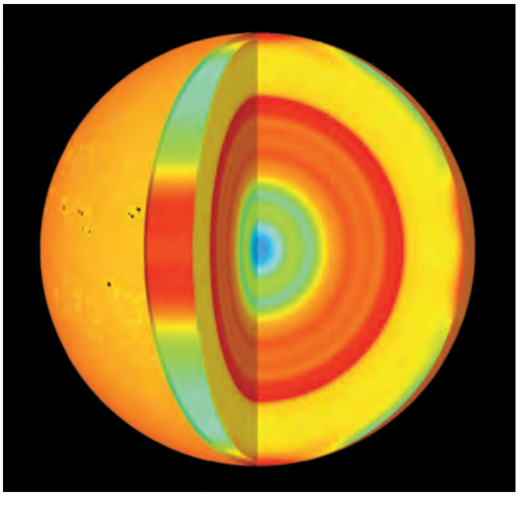


Figure 2.11 Variations of sound speed within the Sun. This figure shows the difference between a solar model and results obtained from the analysis of data from the SOHO mission. Red and blue/green areas show regions in which the sound speed is respectively higher and lower than predicted. The speed of sound is related to temperature and the red and blue/green zones correspond to temperatures that are respectively higher and lower than predicted.
(A. G. Kosovichev, Stanford University)

Helioseismology experiments have been used to provide confirmation that solar models do provide a good description of the Sun. One such illustration of this is shown in Figure 2.11, which shows the speed of sound within the Sun as compared to that predicted by a solar model. The speed of sound through a gas depends on its temperature, hence differences in sound speed can be used to trace differences in temperature. In Figure 2.11 the areas shown in red are hotter than predicted by the solar model, and blue areas are cooler than expected. The first point to note is that the deviation from the solar model that has been adopted here is very small; throughout most of the Sun the difference between predicted and measured sound speed is less than 0.7%. So the solar model appears to be a good description of reality. Furthermore, these results confirm that the core temperature of the Sun is 1.56×10^7 K. The deviations from the solar model are also of interest, as they highlight unexpected features within the Sun. The boundary between the radiative and convective zones at $0.68R_\odot$ is prominent; the temperature in the radiative zone is higher than expected as this boundary is approached, while the temperature in the convective zone is the same as the solar model predicts. This slight temperature enhancement in the radiative zone is an unexpected result, and one which has led to refinement and improvement of the solar model.

However, the overall conclusion from helioseismology studies has been a confirmation that solar models provide a very good description of the structure of the Sun.

Helioseismology has also been used to investigate the rotational motion of material within the Sun. We have already seen from the study of sunspots that the Sun rotates differentially at its surface and helioseismology provides a method of measuring how this pattern of rotation changes with distance into the Sun. The result of one such study using data from SOHO is shown in Figure 2.12 which illustrates that the pattern of differential rotation that is seen at the surface extends throughout the convective zone, but below this, there is an abrupt change to rotation that follows the same behaviour as that of a solid body.

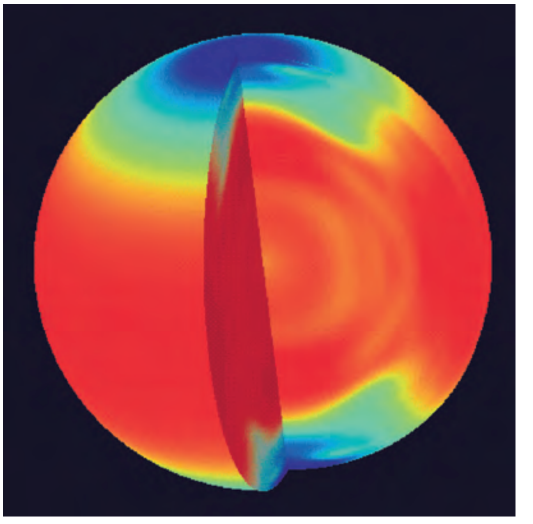


Figure 2.12 The rotation of the solar interior as determined from helioseismological data. The figure shows the rate of rotation: regions that are the same colour have the same period of rotation. The colour coding is such that the red regions are rotating faster than average, the yellow regions are rotating slower than average, and the blue regions have a yet slower rotation rate. (Fleck *et al*., 2000)

Observations of solar neutrinos

Although electromagnetic radiation has a hard time escaping from the Sun's core, the neutrinos produced there have no such difficulty. Observations of **solar neutrinos** potentially provide a very direct test of our ideas about solar nuclear reactions but, unfortunately, the low interaction rate that allows the neutrinos to escape from the Sun also makes them very hard to detect when they reach (and pass through) the Earth. Nonetheless, experiments to monitor solar neutrinos have been running since 1970. The first of these experiments was set up in the Homestake gold mine in South Dakota, USA. (The underground location has no influence on the neutrinos, but it does cut out various other particles that might otherwise lead to spurious results.) This neutrino detector (Figure 2.13) is rather unusual; it consists of a large tank containing 610 tonnes of perchloroethylene (tetrachloroethene, C_2Cl_4) – a liquid used in the dry-cleaning business. As neutrinos flood through the tank, one occasionally interacts with a chlorine nucleus to produce a ${}^{37}_{18}Ar$ nucleus. At the end of a typical 80 day run the tank is emptied, its contents are analysed, and the ${}^{37}_{18}Ar$ nuclei are counted. This is no mean feat, because usually only about 50 argon nuclei are found amongst the roughly 10^{31} nuclei in the tank.

Figure 2.13 The tank of C_2Cl_4 used to trap solar neutrinos at the Homestake gold mine in South Dakota. (Brookhaven National Laboratory)

The results of the Homestake mine experiment have been a source of controversy for many years. The small number of observed neutrinos implies a rate of neutrino production that is only about one-third of that expected. This mismatch between observation and theory is referred to as the **solar neutrino problem**. Since the first results were published from the Homestake mine experiment, the deficit of solar neutrinos has been confirmed by other experiments that detect solar neutrinos in a variety of ways. One of the problems faced by researchers was that the early neutrino detection techniques, such as that based on C_2Cl_4, were sensitive to only the most energetic solar neutrinos, and those come from a relatively rare process involving the decay of boron (${}^{8}_{5}B$) nuclei. It was not until the mid-1990s that experiments confirmed the Homestake mine result by measurement of the neutrinos produced from the ppI chain itself.

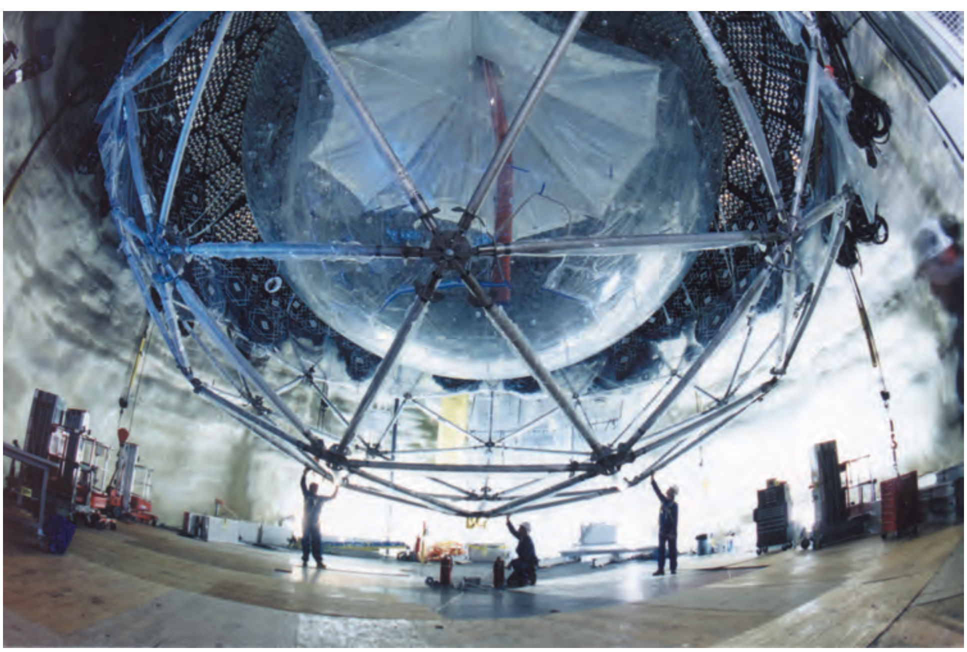
Up until the late 1990s opinion was divided as to what could be the most likely mechanism to produce the observed deficit. One view was that solar models were flawed in some important respect; for instance that it may have been incorrectly assumed that there is no mixing of material in the core. This view began to lose support as results from helioseismology experiments started to show that solar models were valid. The most popular alternative explanation was that the nature of the neutrino itself causes the problem. Physicists know of three kinds of neutrino, respectively referred to as 'electron neutrino' ν_e, 'muon neutrino' ν_μ and 'tauon neutrino' ν_τ. The neutrinos created in the Sun's core should all be electron neutrinos and this is the only sort of neutrino that can be detected in the tank of C_2Cl_4. As a solution to the solar neutrino problem, it was suggested that a particular kind of interaction, between the neutrinos leaving the core and the solar material through which they must pass, causes the neutrinos to change type – with the consequence that only about a third of the neutrinos emerging from the Sun are still electron neutrinos.

The resolution of the solar neutrino problem was one of the most important issues facing solar astronomy at the end of the 20th century, and in order to settle the matter, several large solar neutrino experiments were commissioned. One such experiment is the Sudbury Neutrino Observatory (SNO, Figure 2.14a) in Canada, which like the Homestake mine experiment, is located deep underground. The detector in this case is a spherical acrylic tank that holds 1000 tonnes of *heavy water* (Figure 2.14b). The tank is surrounded by 9500 photosensitive cells that detect flashes of visible light that result from neutrino interactions within the tank. There are several types of interaction that neutrinos may have with heavy water. A molecule of heavy water is identical to normal water except that the hydrogen (${}^{1}_{1}H$) atoms are replaced by atoms of deuterium (${}^{2}_{1}H$). A deuterium nucleus may interact with neutrinos in several ways. In one reaction, an electron neutrino can interact with the neutron in the deuterium nucleus, resulting in the production of a proton and a high-energy electron. As this electron moves at very high speed away from the site of reaction, it produces a flash of light that can be detected by the photocells. However, another reaction, in which the deuterium nucleus is essentially split into a proton and a neutron, can be triggered by any type of neutrino – muon neutrinos and tauon neutrinos, as well as electron neutrinos. This reaction causes the emission of a γ-ray as the neutron is absorbed by another nucleus within the tank. The γ-ray interacts with an electron causing it to recoil at high speed, and again a flash of visible light is produced. The flashes of light from different types of interaction differ from one another and it is possible to determine which reaction has given rise to an observed flash. The advantage of SNO over earlier experiments is that it provides a way to measure the flux of all three types of neutrino rather than just the electron neutrinos. This capability allows detailed

testing of the hypothesis that neutrinos change their type. In 2001 the team running SNO announced their first results, which indicated that neutrinos do indeed change type. This result supports the view from helioseismology that solar models are a good description of the Sun, and provides an exciting new challenge to theoretical particle physicists to explain why the neutrino should behave in this way.

(a)

Figure 2.14 The Sudbury Neutrino Observatory (SNO). (a) An artist's impression of the deep underground chamber that houses the observatory, and (b) a photograph of the spherical acrylic tank that is filled with heavy water to act as a neutrino detector. (SNO)



(b)

QUESTION 2.6

Estimate the number of solar neutrinos per second passing through a $0.01\,\mathrm{m}^2$ detector located on the Earth and pointed directly towards the Sun. You may assume that the distance from the Sun to the Earth is 1.50×10^{11} m, and you will find it useful to know that the surface area of a sphere of radius r is $4\pi r^2$. Take care to write down any other assumptions you make in obtaining your estimate.

2.3 Solar activity

In Chapter 1 you saw that the Sun has an 11-year cycle of activity, and that this cycle is clearly shown by the numbers of sunspots. In this section we return to investigate indicators of solar activity in more detail. We start by looking again at sunspots, and in particular, at the relationship between sunspots and magnetic fields. We shall see that other phenomena which are related to solar activity are also magnetic in origin. After considering the question of how the most energetic events linked to solar activity are powered, we finally consider the global changes to the Sun's magnetic field that occur over the activity cycle.

2.3.1 Sunspots and active regions

You saw in Section 1.2 that sunspots are a prominent feature of the photosphere, and that the number of sunspots follows an 11-year cycle. An individual sunspot usually shows the features that are clearly visible in the largest sunspot in Figure 2.15. There is a dark central region called the umbra that is surrounded by a somewhat lighter region called the penumbra. The pattern arising from granulation that is seen elsewhere in the photosphere is disrupted in the sunspot. The penumbra characteristically shows fine filaments, which tend to be directed radially away from the centre of the sunspot. (These are quite distinct from the much larger scale chromospheric filaments that were discussed in Section 1.3.1.) The umbra is notable for the complete absence of granulation and the low incidence of other features, although occasionally bright 'dots' or filamentary structures are seen within this region.

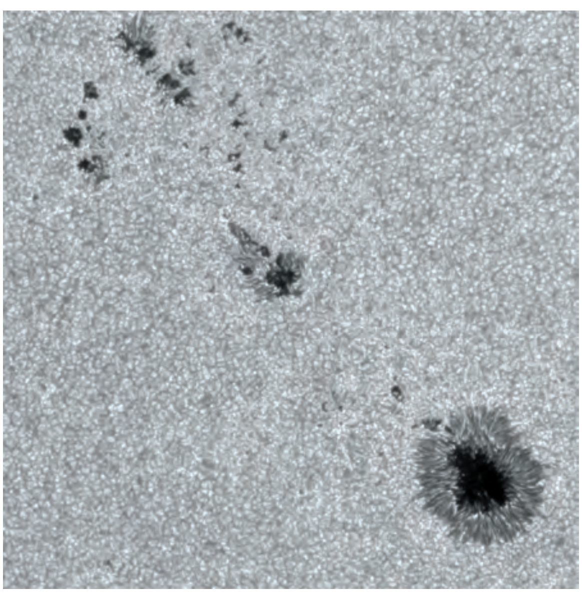


Figure 2.15 A group of sunspots observed on 12 May 1998. The extent of this image is approximately 270 000 km × 270 000 km. (Royal Swedish Academy of Sciences)

Measurements of magnetic fields reveal that all sunspots are characterized by magnetic field strengths that are much higher than elsewhere in the photosphere. While it seems clear that the properties of sunspots arise from these strong magnetic fields, the detailed physical processes are not well understood. One currently favoured idea is that the intense magnetic field suppresses convection. This reduces the rate of energy transport to the photosphere and this results in the relatively cool regions which characterize the sunspot.

In order to describe the way in which a magnetic field varies within a sunspot, or more generally in any region, it is useful to use the concept of magnetic field lines as described in Box 2.2.

BOX 2.2 MAGNETIC FIELD LINES

We saw in Box 1.1 that the effect that a magnet has throughout a region of space can be described by using the concept of the magnetic field. The magnetic field describes both the direction and magnitude of the magnetic force at all locations in a region of space. We also saw from Box 1.1 how a magnetic field can be represented by arrows that show the direction and strength of the field. This idea can be extended by imagining the path that would be drawn in space if we were to follow one of these arrows for a short distance, and then to consider the arrow at the new position. This path is called a field line, and provides a good way of visualizing how a magnetic field varies throughout a region of space. If you look at Figure 2.16a and join the arrows that seem to be connected, this will give a good approximation to the field lines as shown in Figure 2.16b.

The interpretation of **magnetic field lines** is that their direction at any point in space is simply the direction in which a compass needle would orient itself. Furthermore the direction of the field is indicated by small arrows on the lines. By convention, a magnetic field such as that shown in Figure 2.16b is considered to be directed *from* a north magnetic pole *to* a south magnetic pole. The strength of the magnetic field is interpreted not by the length of any arrows, but by how closely together the magnetic field lines are packed.

Note also that magnetic field lines form continuous closed loops: they do not start or end at any point in space. Although it may appear from a diagram such as Figure 2.16b that the field lines start and end at the boundary of the magnet, in fact, the field lines pass through the magnet such that every line is actually part of a continuous loop.

The magnetic field that is produced by a bar magnet is special because many magnetic fields in nature show a similar pattern. This field is called a **dipole field** because it has two poles, one north and one south, and has a pattern as shown in Figure 2.17. The field lines emerge from the north pole of the magnet and loop around to the south pole. The magnetic field of the Earth follows this pattern well in the region that is close to the surface of the planet. Rather confusingly, the magnetic pole in the Earth's *Northern* Hemisphere is actually a *south* magnetic pole.

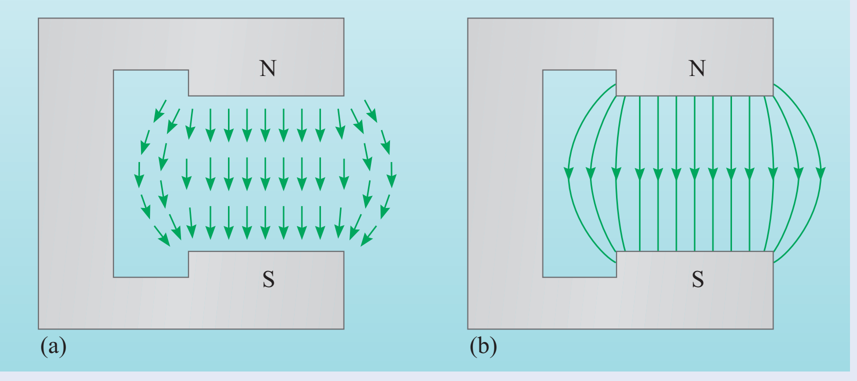


Figure 2.16 (a) The magnetic field represented by the length and direction of arrows, and (b) the magnetic field pattern of the same magnet.

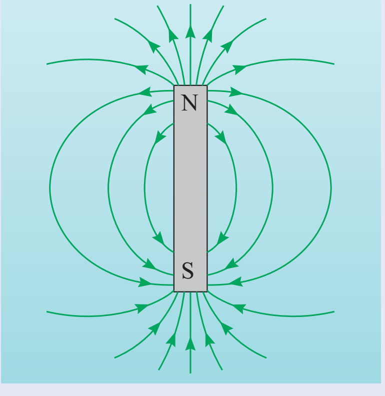


Figure 2.17 The dipole field of a bar magnet.

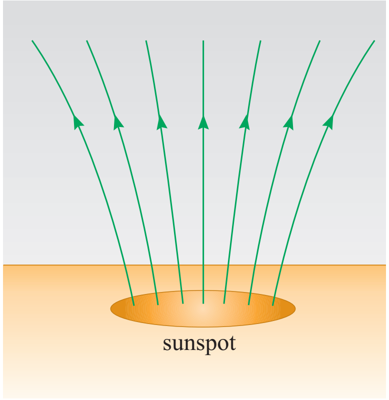


Figure 2.18 Magnetic field lines in a sunspot. (Note that in an individual sunspot, the field lines may run either into, or out of, the Sun.)

The magnetic field lines from an individual sunspot tend to run almost vertically in or out of the umbra. In the penumbra, the magnetic field curls over as shown in Figure 2.18. Since the field can be directed either upwards or downwards, an individual sunspot will resemble either a north or south pole, and these two possibilities are referred to as north (or positive) and south (or negative) **polarities** respectively.

- ■ A sunspot is observed with field lines in the umbra that are running into the Sun. What is the polarity of the sunspot?
- ❑ Since magnetic field lines run into the sunspot, it is similar to a south pole and so has south (or negative) polarity.

There are no differences in structure or behaviour between sunspots of different polarities, but the observed polarities of sunspots do provide an important clue as to the nature of the solar cycle.

It was noted earlier that sunspots often exist in groups, and in particular, there is a tendency for them to occur in pairs. When the magnetic fields of these sunspot pairs are measured it is always found that the two members of the pair have opposite polarities. A schematic illustration of the field lines associated with a pair of sunspots is shown in Figure 2.19. The field lines emanate from one sunspot and form a loop which arches above the photosphere and returns to enter the Sun at the other member of the pair. This suggests that the formation of a pair of sunspots is due to a 'bundle' of magnetic field lines breaking out from the Sun's interior. We will investigate how the magnetic field of the Sun may give rise to such a pattern when we consider the solar cycle in Section 2.3.5.

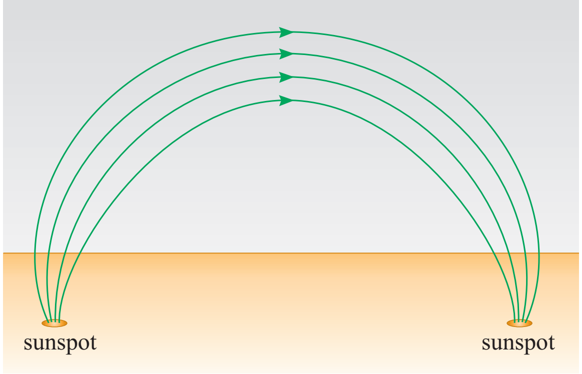


Figure 2.19 The loop of magnetic field lines that connect two members of a pair of sunspots.

Sunspots are clearly associated with very intense magnetic fields, and the sunspots themselves are an indicator of solar activity. However, when the magnetic field over the photosphere is measured, it is found that extended areas around sunspots show the same polarity patterns as the sunspots, but with lower field strengths. The strength of the magnetic field in the photosphere can be mapped because of a phenomenon (called the *Zeeman effect*) that arises when atoms or ions are subject to a strong magnetic field. The electronic energy levels in atoms and ions are altered by magnetic fields, and this has the effect that individual spectral lines split into two or three lines of slightly different wavelengths. The size of this effect is small; for example, in the strong field of a sunspot the separation between split spectral lines may only be about 10^{-3} nm. Such splitting can be measured using a specialized instrument called a **magnetograph** that can map the magnetic field strength over the visible surface of the Sun. A map produced in this way is called a **magnetogram**.

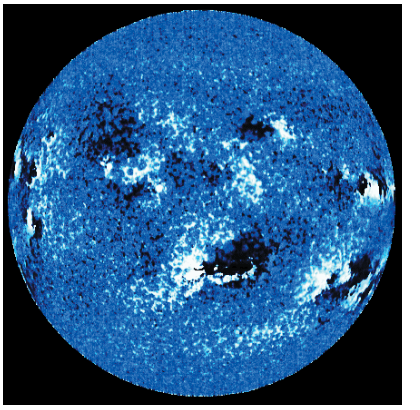
A magnetogram that was obtained at a time of high solar activity is shown in Figure 2.20. The strong fields associated with sunspots are clearly visible, as are the surrounding regions that show the same polarity as the sunspots, but have moderate field strength. In addition there are extended regions which show moderate field strengths but are not associated with any sunspots. All of these regions of enhanced magnetic field are locations in which the various phenomena that are associated with solar activity are likely to occur, and hence are called **active regions**.

- List the phenomena associated with solar activity that you have already encountered.
- Sunspots (Section 1.2.2); plages (Section 1.3.1); filaments and prominences (Section 1.3.1); solar flares (Section 1.4.3).

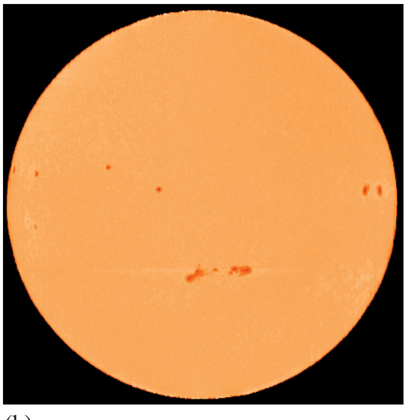
All of these phenomena can be understood to some degree in terms of the behaviour of the underlying magnetic field. It has already been noted that the relatively low temperature of sunspots probably arises from a suppression of photospheric convection by the enhanced magnetic field. Plages are bright regions of the chromosphere, which occur in approximately the same location as bright photospheric regions called **faculae**. Both plages and faculae seem to be related to sunspots, since they appear in active regions before the emergence of a sunspot (or sunspot group) and remain visible until after any sunspots have disappeared. Hence plages and faculae are also clearly magnetic in origin, but the exact mechanisms that are at work are not well understood.

You saw in Section 1.3.1 that the terms *filament* and *prominence* refer to the same physical phenomenon – that of a strand of material that is supported above the photosphere. In fact, filaments and prominences can occur both at times of solar activity and quiescence. The type of filaments that are seen in active regions tend to be much more variable than those in quiescent parts of the Sun. Active prominences and filaments are fairly short-lived, maybe only existing for days or even just a few hours, compared to quiescent filaments that may be stable for weeks or months.

Another feature that is very commonly seen in active regions is called a **coronal loop** (Figure 2.21). These are typically seen in ultraviolet and X-ray images of the Sun and take the form of closed loops that contain plasma at temperatures that may exceed 10^6 K. As their name suggests, these features extend up into the solar corona. The shape of coronal loops suggests an association with the magnetic field within the active region.



(a)



(b)

Figure 2.20 Simultaneous solar measurements of magnetic field and photospheric emission.
(a) The upper panel of this image shows a magnetogram of the solar photosphere. The colour coding of the magnetogram is: black – negative field; blue – weak field of either polarity; white – positive field.
(b) The lower panel shows the intensity of the photosphere over a wide range of visible wavelengths, and hence shows the location of sunspots. (Data provided by National Solar Observatory)

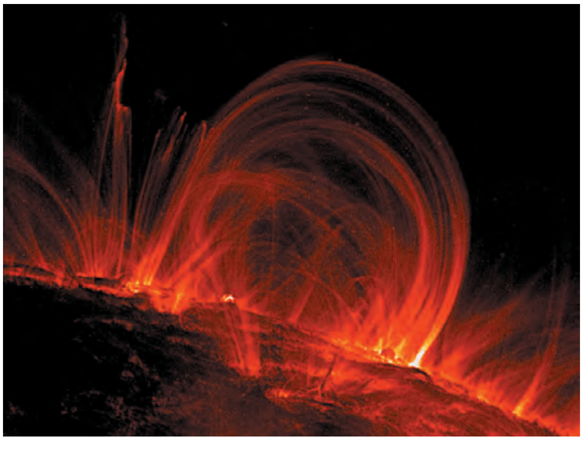
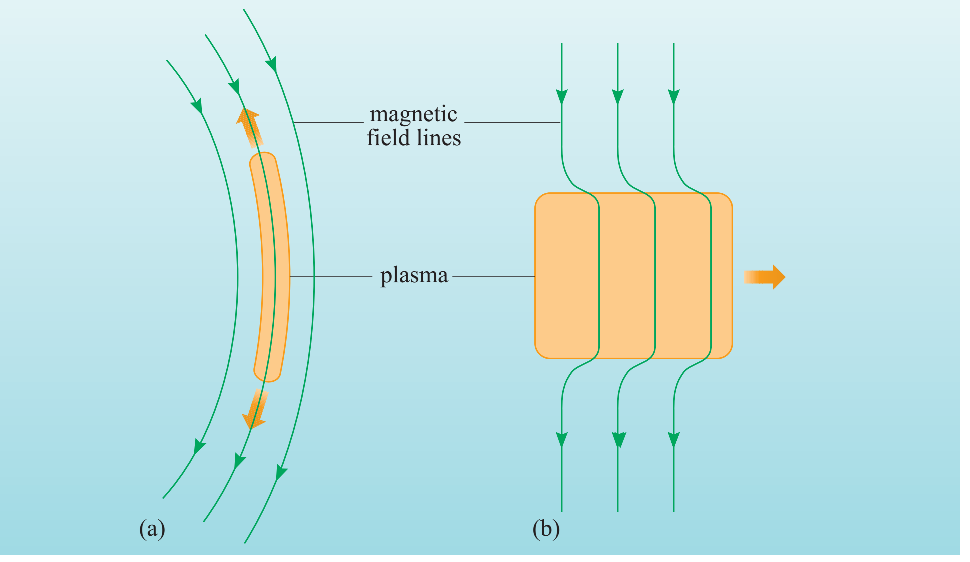


Figure 2.21 A coronal loop. This image shows the emission at 17.1 nm – this is in a part of the electromagnetic spectrum that lies at the short wavelength end of the ultraviolet band. This image was obtained from a telescope on the TRACE (Transition Region And Coronal Explorer) satellite. (TRACE (NASA))

- Compare the image of a coronal loop (Figure 2.21) with the field lines associated with a pair of sunspots (Figure 2.19). What does this suggest about material within the prominence?
- The shape of the active prominence seems to follow loops of the magnetic field lines. This suggests that material in the prominence can only move along magnetic field lines.



This is indeed what happens. The material within a coronal loop is a plasma, and an important property of plasmas is that they interact very strongly with magnetic fields. In the presence of a sufficiently strong magnetic field, plasma is constrained to move along the magnetic field lines. Hence an ionized gas in a strong magnetic field may move along field lines, but cannot escape from the closed loops that field lines form. This then explains the appearance of the coronal loop: it is a loop of field lines that has trapped the plasma.

Figure 2.22 The interaction between a plasma and a magnetic field: (a) in cases where the magnetic field is strong the plasma is constrained to move along magnetic field lines; (b) when the magnetic field is relatively weak, magnetic field lines are carried with the flow of plasma.

In passing, it is also worth noting that the interaction between magnetic fields and plasmas may take on a somewhat different character when the field is relatively weak, but the flow of plasma is relatively strong. In this case the plasma will carry the magnetic field lines with it. The magnetic field in such a situation is often referred to as being 'frozen-in' to the plasma, a description that embodies the idea that the field lines have to follow the motion of the plasma. The two extremes of behaviour – a plasma that is confined to a strong magnetic field, and a magnetic field that is carried with a strong flow of plasma – are illustrated schematically in Figure 2.22.

The features of solar activity that we have seen so far, sunspots, plages, active prominences and loops, all depend on the existence of strong magnetic fields in active regions. In the following two sections we will look at phenomena that are thought to arise when those magnetic fields undergo very rapid changes.

2.3.2 Solar flares

Solar flares were mentioned briefly in Section 1.4 as being rapid and energetic outbursts of electromagnetic radiation. The duration of a flare is typically between about 100 s and a few 1000 s. During that time the flare may release up to about 10^{25} J in radiation, and material in the region of the flare may be heated to temperatures of over 10^7 K. The regions in which flares occur are small: typically less than 0.1% of the solar disc. These regions often lie between sunspot pairs or within sunspot groups, although flares may also occur in active regions in which sunspots are not present. Given their association with sunspots and active regions, it is not surprising to find that the incidence of solar flares follows the solar activity cycle. At the time of solar activity maximum, the rate of occurrence of flares is about ten times that at solar minimum.

During a flare, electromagnetic radiation is emitted over a wide range of wavelengths, but the dominant emission is in the X-ray and extreme ultraviolet (a region between the ultraviolet and the X-ray parts of the electromagnetic spectrum). This is illustrated in Figure 2.23 which shows an image taken with an X-ray telescope; the flaring region is far brighter than other emitting regions. Flares are also conspicuous in the γ-ray and the radio parts of the spectrum. At visible wavelengths the emission is usually swamped by the brightness of the photosphere, and only the most energetic bursts can be detected as increases to the broadband

optical flux density. However, solar flares cause a dramatic increase in emission from spectral lines such as Hα, and so can be detected in images that are tuned to the appropriate line wavelength. The emission at different wavelengths varies over the duration of the flare. The first sign of a flare is a rapid, but short-lived outburst that is seen in very energetic X-rays (termed 'hard' X-rays) and in microwave emission. As the outburst dies away in these wavebands, emission in lower energy, or 'soft', X-rays, extreme ultraviolet light and in the Hα line increases and reaches a maximum after a few minutes and then slowly decays away.

The association between solar flares and sunspot groups suggests that flares are essentially a magnetic phenomenon. This is confirmed by observations such as the extreme-ultraviolet image of a solar flare shown in Figure 2.24. The emission comes from hot plasma (at a temperature of about 10^6 K) that is contained within the loops of the magnetic field. Detailed observations of the onset of flares show that the initial burst of hard X-rays often occurs at both foot-points of a magnetic loop. The simplest interpretation of this is that some process high in the loop gives rise to an energetic burst of particles (electrons or protons) which follow the field lines down towards the solar surface. It is the interaction of these particles with material in the chromosphere that gives rise to hard X-ray emission. This interaction also heats the chromospheric material to temperatures exceeding 10^6 K. This hot plasma then rises back up along the loop, probably filling it completely (as seen in the example shown in Figure 2.24). It is this material that gives rise to the lower energy X-ray and extreme ultraviolet emission. This scenario also explains why the soft X-rays and extreme ultraviolet light are emitted at a later time than the initial hard X-ray burst. This sequence of events is summarized in Figure 2.25. It should be stressed however, that this pattern of events is not seen in all solar flares, and that the study of these energetic outbursts is an on-going area of research.

This is only a partial model for a solar flare: it does not address the fundamental problem of how such a rapid outburst of energy can be generated. We will return to this problem later in Section 2.3.4, after considering a different phenomenon that also involves the rapid release of a large amount of energy.

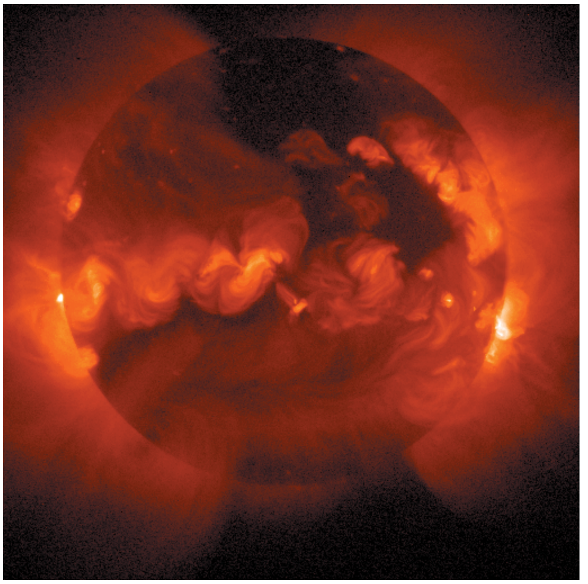


Figure 2.23 An image of the Sun taken with an X-ray telescope in which a solar flare is clearly visible on the western limb (right-hand side) of the Sun. (Yohkoh (ISAS))

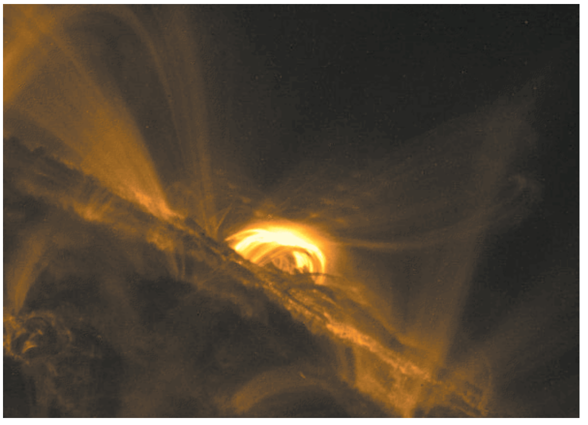


Figure 2.24 An image of a solar flare taken at a wavelength of 17.1 nm (from the TRACE mission). (TRACE (NASA))

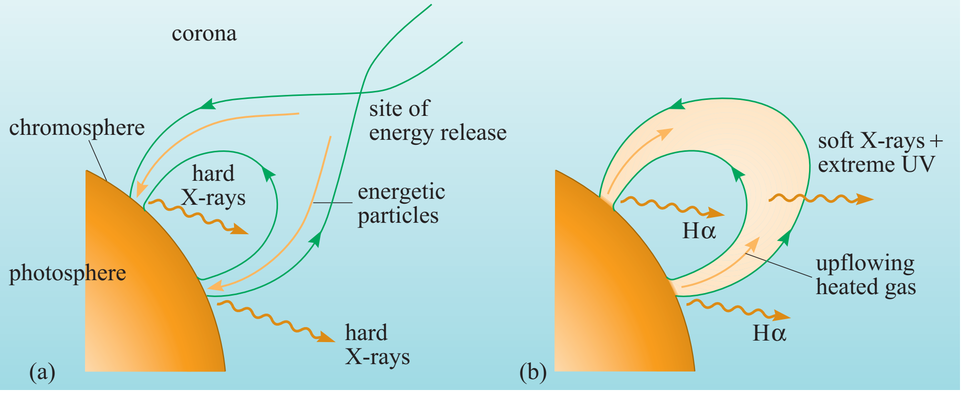


Figure 2.25 The sequence of events in a solar flare. In the first stage (a) an energetic event gives rise to streams of particles that travel down the magnetic field lines and result in hard X-ray emission. This is followed by (b) the upflow of heated material from the chromosphere giving rise to soft X-ray, ultraviolet and Hα emission from the magnetic loop. (Adapted from Lang, 2001)

2.3.3 Coronal mass ejections

We have already seen that the solar corona is a dynamic environment in which large-scale changes can occur very rapidly. While solar flares are rapid outbursts of energy that are characterized by intense bursts of electromagnetic radiation, there is a different type of event, called a coronal mass ejection (CME), that manifests itself as a violent outflow of coronal material. Some of the best observations of CMEs have been obtained using the coronagraph on board the SOHO spacecraft that is able to view the corona over a range of distances from $1.1R_\odot$ to about $30R_\odot$ from the solar centre. Figure 2.26 shows a sequence of SOHO observations that illustrate the evolution of a typical CME over an interval of about eight hours. Frame (a) shows the corona prior to the start of the coronal mass ejection. The CME itself appears in frame (b) as a 'bubble' projecting towards the top left-hand corner of the frame. Through subsequent frames this bubble expands outwards from the surface of the Sun, reaching a size of tens of $R_\odot$ in a matter of hours. After several hours the CME passes out of the field of view of the coronagraph, but continues to travel out into the interplanetary medium.

QUESTION 2.7

From the sequence of images shown in Figure 2.26, estimate (a) the speed of the front of this CME, and (b) the time that it would take for this CME to travel a distance equal to the Earth's distance from the Sun. What assumptions do you need to make in carrying out your calculations?

The mass of ejected material in a large CME event is typically between 5×10^{12} kg and 5×10^{13} kg, and CME events occur with an average frequency of about one per day. The rate of occurrence depends on the level of solar activity; at solar maximum there are about three CMEs per day, while at solar minimum the rate is about one-tenth of this.

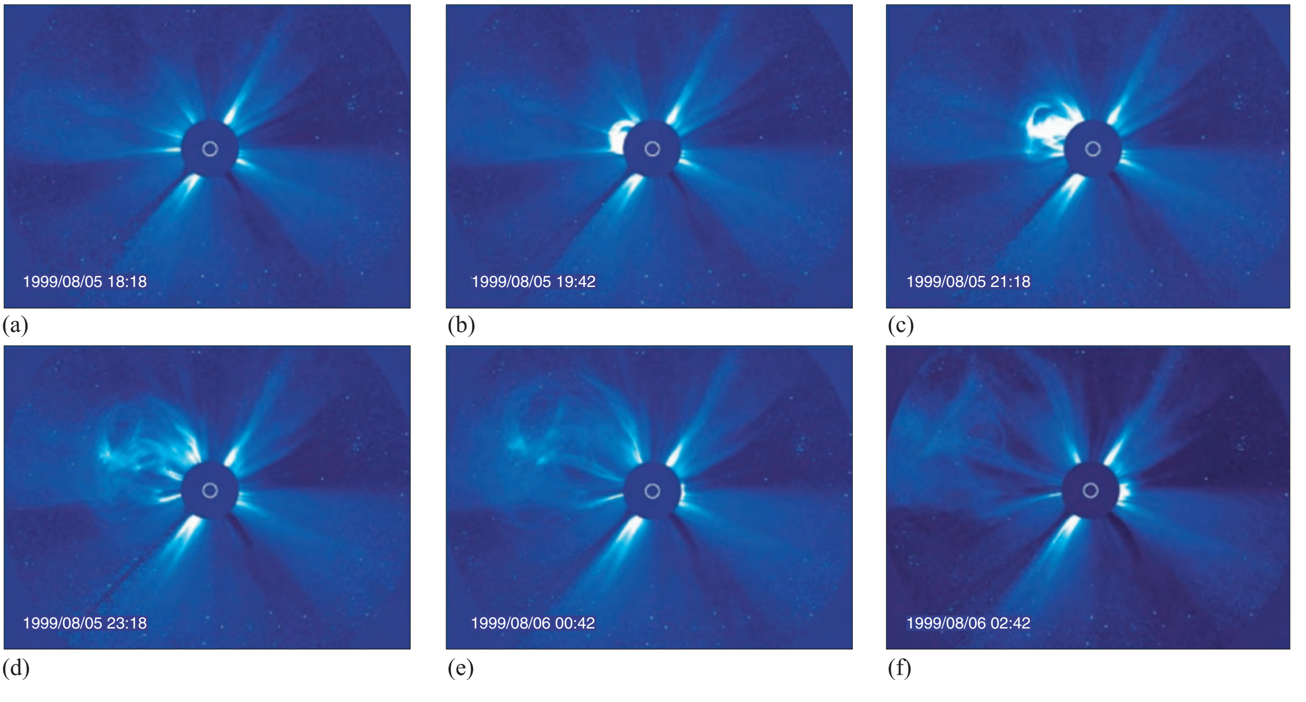


Figure 2.26 The evolution of a coronal mass ejection observed over an 8-hour interval on 5–6 August 1999. The dark disc blocks the Sun so that the coronagraph can observe structures in the corona in visible light. The white circle represents the size and position of the Sun. (Fleck *et al*., 2000)

The loop-like shape of CMEs, such as that shown in Figure 2.26, suggest that the plasma is confined within a magnetic field. The essential feature here is that the magnetic field is moving and plasma is being forced to move with the field. The origin of these ejection events probably lies in a rapid reconfiguration of the magnetic field in the lower parts of the solar corona. This allows energy that is stored in the magnetic field to be released suddenly, causing a violent outflow of coronal material.

QUESTION 2.8

Calculate the total kinetic energy of a large coronal mass ejection. Use the estimated speed from your answer to Question 2.7 and assume that the ejection has a total mass of 5×10^{13} kg. Express your answer as an order of magnitude estimate (i.e. to the closest power of ten).

2.3.4 Magnetic reconnection

Solar flares and coronal mass ejections both require a mechanism for the sudden release of energy, and it is also clear that both phenomena are magnetic in origin. Rather than concentrating on the features that distinguish these two types of event, solar physicists take the approach that there is likely to be a single underlying mechanism that drives both solar flares and CMEs. The starting point for such theories is the idea that magnetic fields act as a store of energy. It is not too difficult to convince yourself of this; if you take two bar magnets and arrange them end on such that like poles are next to one another, you will find that you have to push the magnets together to get the poles to touch. If you release one magnet, some of the energy that you supplied by pushing the magnets together, will be returned to that magnet in the form of kinetic energy as it is repulsed. It is the interaction between the magnets that allows energy to be stored. This interaction is described by the magnetic field: as the magnets are pushed together the field pattern changes and stores energy that can be recovered when one of the magnets is released.

The energy stored by magnetic fields that are found in the solar corona can easily meet the energy requirements of solar flares and CMEs. The problem, which is not fully solved, is how the magnetic field energy can be released rapidly enough to explain the violent outbursts seen in solar flares. Despite this problem it seems likely that a process called **magnetic reconnection** is at play. The fundamental idea is that in some regions, field lines that are running in opposite directions are squeezed together as shown in Figure 2.27. The region that separates field lines that run in opposite directions has a particularly important role. Remember that all solar magnetic fields exist in a medium which is ionized – a plasma. Because they contain electrons and ions that are not bound together as neutral atoms, plasmas are able to conduct electricity. In fact, the existence of the magnetic field causes an electric current to run along the boundary that separates regions in which the field runs in opposite directions. In Figure 2.27 the electric current would, in fact, run out of the page, perpendicular to the field lines. The flow of an electrical current through a conductor usually gives rise to heating; this is the principle behind such everyday appliances as kettles and electric heaters. Similarly, the current flowing through the plasma also causes heating. Heat is generated at the boundary between the different field directions.

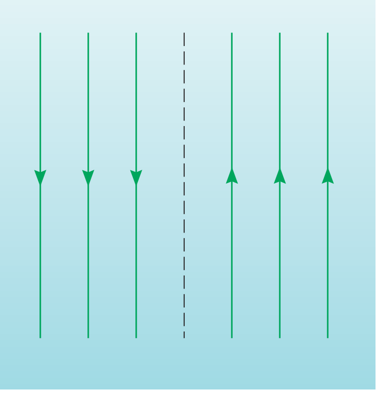


Figure 2.27 The configuration of magnetic field lines in a region where magnetic reconnection can occur. The dashed line shows the boundary between regions of opposite polarity.

- ■ What is the source of energy for this heating?
- ❑ The energy comes from the magnetic field – this causes the electrical current that results in heating.

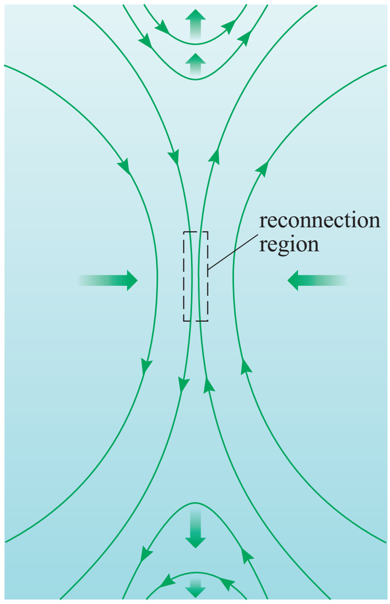


Figure 2.28 The configuration of magnetic field lines if two regions of opposite polarity are pushed together. Note that reconnection only occurs over a region of rather limited extent.

Since energy is removed from the magnetic field, the magnetic field must drop in strength as heat is produced. In terms of Figure 2.27, field lines of opposite direction essentially cancel each other out. However, this figure does not account for the fact that all magnetic field lines are actually in the form of continuous loops: in reality there can only be certain parts of field lines where they come together as shown. If we consider a somewhat more realistic situation in which two regions come together in the way shown in Figure 2.28, then there will be heating in the region where the field lines are closest. Two opposing field lines may disappear from this region, but at the edge of the region, the lines connect up to one another to form a single field line. This field line behaves in a way that is similar to an elastic band under tension. Initially, the field line has a very sharp bend in it at the point where the reconnection took place, but immediately tries to straighten itself out, and in doing so it moves rapidly away from the reconnection region. Because plasmas are forced to move with the field lines, this rapid motion of the field lines can accelerate particles in the plasma to high velocities. Hence the process of reconnection can result in the conversion of energy stored in the magnetic field into the kinetic energy of particles.

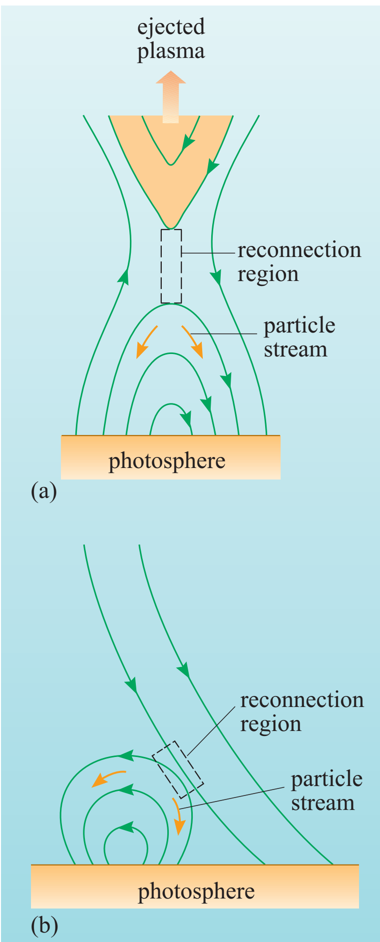


Figure 2.29 Two scenarios that have been proposed for the magnetic field configuration that may cause solar flares and coronal mass ejections. While both scenarios account for the observed features of solar flares, the ejection of a large mass of plasma is more likely in case (a) than in case (b). ((a) Sturrock, 1980; (b) Heyvaerts *et al*., 1977)

The way in which magnetic reconnection events may give rise to solar flares or CMEs is illustrated in Figure 2.29, which shows configurations of magnetic fields that might explain these events. In both scenarios that are illustrated, a common feature is the fact that the reconnection event occurs high in a coronal loop. This results in particles being accelerated down to the foot-points of the loops to give the characteristic outbursts seen as solar flares. In the case shown in Figure 2.29a, the field reconfiguration after reconnection is such that it can drive the ejection of any plasma that lies above the reconnection point. In the case shown in Figure 2.29b, there is a somewhat different magnetic field configuration; the reconnection event causes a solar flare, but results in a smaller ejection of plasma. It should be noted that these are speculative ideas about the way in which reconnection takes place. Unfortunately, at present, it is not possible to map the magnetic fields in the corona in a way that would allow these ideas to be tested.

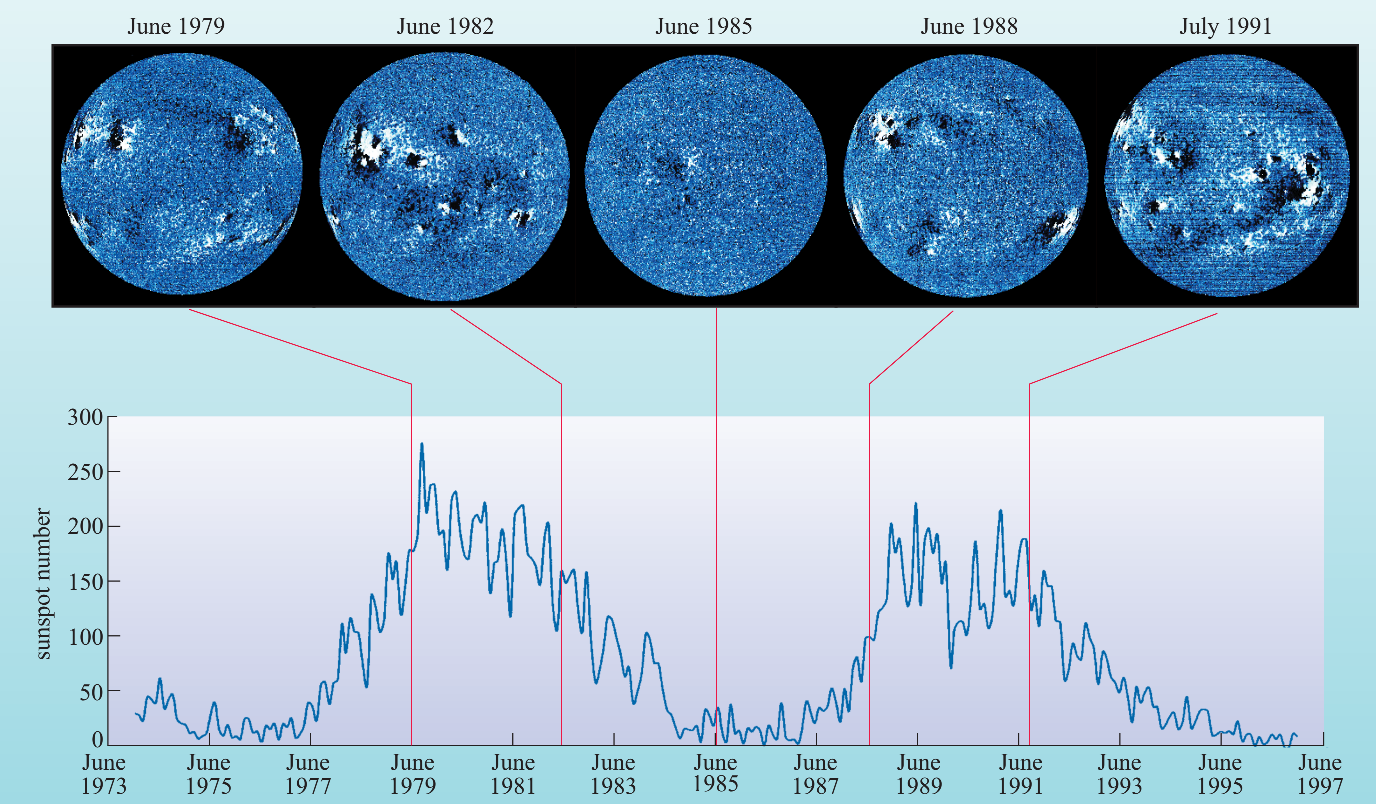
Even so, magnetic reconnection looks like a promising mechanism to provide the energy for solar flares and CMEs. It should be noted however, that there are many problems to be solved before a full understanding of the reconnection process is achieved. In particular, the sort of scheme that has been outlined above does not seem to release energy quickly enough to explain the observed behaviour of solar flares. This area thus remains a topic of great interest to solar physicists.

2.3.5 The solar cycle

In Chapter 1 we saw that the number of sunspots, and all phenomena associated with solar activity, undergo cyclic variability with a period of about 11 years. This recurrence of solar activity is termed the **solar cycle**. So far in this chapter we have considered how *local* variations in the solar magnetic field give rise to phenomena related to solar activity. To gain some understanding of the solar cycle, we need to consider the *global* behaviour of the solar magnetic field, i.e. the way in which the field behaves over the entire Sun.

Figure 2.30 shows a series of magnetograms taken at different times during the solar cycle. There are several important observations that can be drawn from this sequence. The first is that, as might be expected, the occurrence of local enhancements of the magnetic field depends very strongly on the solar cycle: at solar maximum there are many regions of increased magnetic field, at solar minimum these regions are scarce. These areas correspond to active regions on the Sun. A second observation relates to the polarity of sunspot pairs. You have already seen that sunspots tend to be formed in pairs of opposite polarity. These pairs tend to be aligned roughly along lines of constant latitude, but the sunspot that is further ahead in the sense of the rotation of the Sun (called the leading sunspot) is closer to the equator than the trailing sunspot. This behaviour also applies to active regions as a whole; they also show this bipolar behaviour and orientation with respect to the Sun's equator. At the start of the solar cycle, when active regions first start to appear after solar minimum, the location of active regions tends to be at high latitudes, both north and south of the equator. As the solar cycle progresses, the band in which active regions occur migrates towards the equator.

Figure 2.30 The upper panel shows a series of magnetograms taken over a 22-year period, while the lower panel shows a quantity (called the sunspot number) that measures the number and extent of sunspots. The colour coding of magnetograms is: black – negative field; blue – weak field of either polarity; white – positive field. (Data provided by National Solar Observatory)



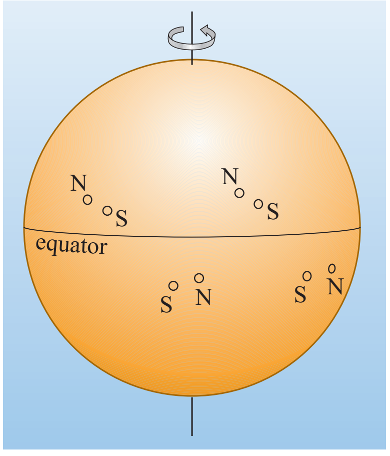


Figure 2.31 Between two solar minima, the polarity of sunspot pairs is such that one polarity always leads in one hemisphere and the opposite polarity leads in the other hemisphere.

A remarkable feature of the polarities of sunspots within a sunspot pair is that at a given time, in one hemisphere, the polarity of the leading sunspot is always the same. In the other hemisphere at the same time, it is the opposite polarity that always leads. This is illustrated schematically in Figure 2.31 and careful inspection of Figure 2.30 should allow you to conclude that this behaviour applies to active regions as well as sunspot pairs. After solar minimum, when sunspot pairs are first seen in a new cycle, the pattern of polarity switches over in the sense that what was the polarity of leading sunspots in the northern hemisphere now becomes the polarity of leading sunspots in the southern hemisphere and vice versa.

■ What is the period of the solar cycle, if it is defined as the time between two solar maxima which have the same behaviour of magnetic polarity?

❏ 22 years. The time between solar maxima is 11 years, but two maxima that are 11 years apart will show opposite polarities for the leading sunspots. Hence it will be another 11 years before the pattern of polarities is repeated.

So when the behaviour of the magnetic field is studied in detail, the solar cycle has a period of 22 years rather than 11 years.

More detailed measurement of the solar magnetic field shows that at the time of solar minimum, the Sun has a global magnetic field that is a reasonable approximation to the dipole field (see Box 2.2). Between minima the field pattern is complicated and characterized by the bipolar active regions. By the time that the dipole field reappears at the following solar minimum, the polarity of the entire field has swapped over.

Solar physicists are yet to reach a full understanding of the mechanism that drives this magnetic cycle. However a scheme which explains some of the broad features of the solar cycle was proposed in 1961 by Horace Babcock, and continues to form the basis on which more sophisticated models are based. The key elements of the scenario are the differential rotation of the Sun and the idea of magnetic reconnection (Section 2.3.4). Babcock's scheme is illustrated schematically in Figure 2.32. At the start of the cycle, at solar minimum, the external field is similar to that of a dipole. Within the Sun, the magnetic field lines run close to the surface, either in or just below the convective layer (Figure 2.32b). These field lines are frozen-in to the plasma within the Sun and so are dragged around by any motion of the plasma. The differential rotation of the surface layers of the Sun leads to a 'winding-up' of the field lines. This has the effect of changing the direction of the field lines from the original pole-to-pole direction to being roughly parallel to the equator (Figure 2.32d). The stretching of the field lines also results in a transfer of some of the energy that the surface layers have due to their motion into energy of the magnetic field. As the field gets wound up even further, loops of field lines start to emerge, and form active regions whose polarities agree with observed behaviour (Figure 2.32e). Towards the end of the cycle it is thought that the active regions that are now close to the equator will start to reconnect across the equator (Figure 2.32f and g). This removes the leading regions and forms loops between trailing regions. In Babcock's scenario, the foot-points of these loops migrate towards respective poles, such that a dipole field is regenerated, but now with an opposite polarity to the field at the start of the cycle. This is then the half-way point in the 22-year cycle, the process repeats with all polarities reversed before returning to the original configuration at the end of the cycle.

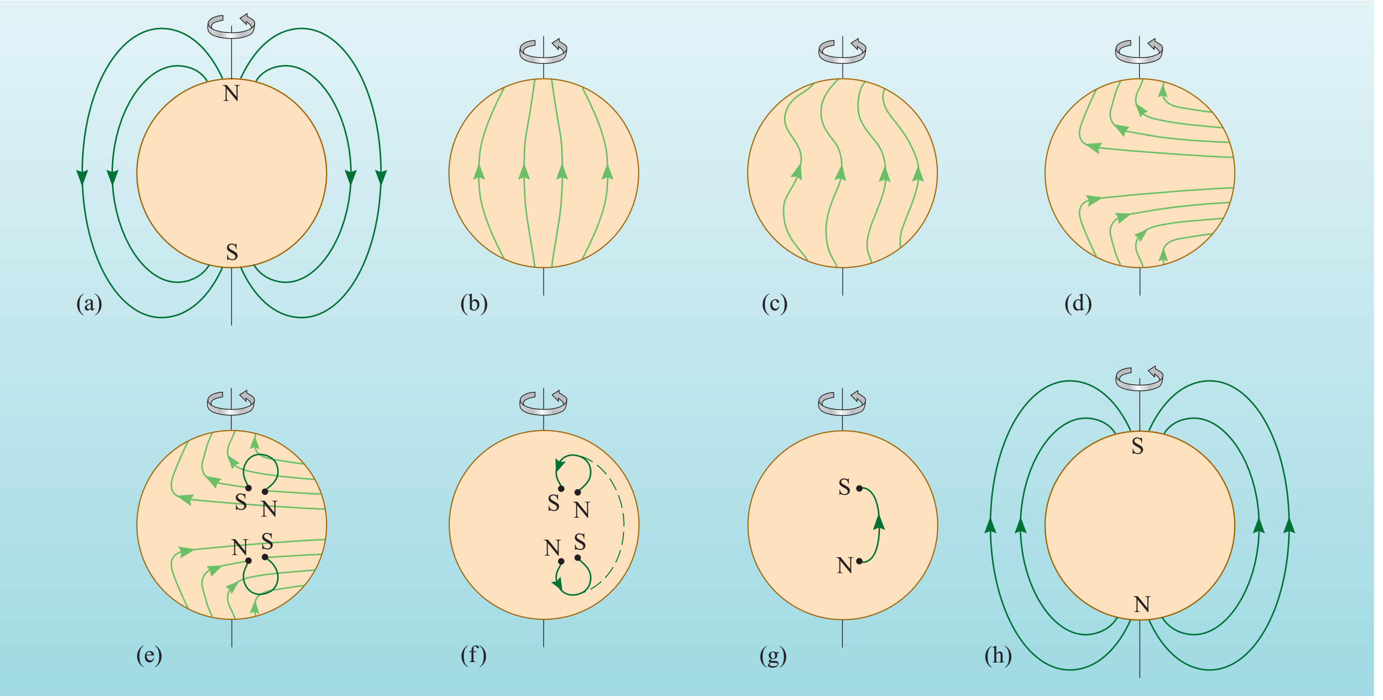


Figure 2.32 A sequence that shows the evolution of the solar magnetic field according to Babcock's scheme. Note that field lines that are interior to the Sun are shown by light green field lines and those exterior to the Sun are shown in dark green. (a) The Sun at solar minimum: the exterior field pattern is close to a dipole field. (b) Inside the Sun these field lines run from one polar region to the other but are close to the surface of the Sun. (c) The interior field lines as the process of differential rotation starts to 'wind-up' the field. (d) As differential rotation continues, the interior field lines become tightly wrapped in bands that run almost parallel to the equator. (e) Loops of field lines emerge to form active regions. (f) As the end of the active part of the cycle approaches, loops reconnect across the equator; leading regions cancel one another out leaving a loop between trailing regions. (g) The new field line now runs approximately in the direction from one pole to another. The foot-points of the new loop now migrate polewards. (h) The result of many such reconnections is a dipole field, but which now has the opposite polarity to that shown in part (a). The cycle then repeats itself but with all polarities swapped.

While this scheme has some very attractive features, in particular that it explains the observed pattern of polarities and orientation of active regions, it is far from being a complete explanation of the Sun's magnetic field. In particular, there is nothing within the model that explains why the Sun should have a magnetic field at all. It is known that magnetic fields can be formed when electrically conducting fluids flow according to certain patterns, and it seems very likely that this sort of effect is responsible for the generation of the magnetic fields. It is however a considerable challenge to theorists to model these sorts of flows within the Sun to reproduce the magnetic field as seen, and this continues to be an active area of research.

2.4 The Sun in space

In the final section of this chapter we take a step back from the Sun to examine some of the effects that it has on its surrounding environment. You have already seen that material is lost episodically from the Sun in coronal mass ejections, and here we will look in more detail at the way that the Sun loses mass continuously by the solar wind. These types of outflow interact with the planets that orbit the Sun and we will briefly look at some of the effects that solar activity can have on the Earth. Finally, we shall investigate how far into space this solar wind extends, and examine the boundary between the environment of the Sun and interstellar space.

Before discussing this in detail, it is useful to sketch out the typical distances that are involved in describing the environment of the Sun. We have already seen that the radius of the Sun ($R_\odot$) is 6.96×10^8 m, and that the solar corona extends to at least several solar radii away from the Sun. The most obvious features in the environment of the Sun are the planets. We will not discuss the planetary system in any detail here, but note that the average distances of planets from the Sun range from 5.79×10^{10} m (Mercury) to 5.90×10^{12} m (Pluto). The average distance from the Sun to the Earth is 1.50×10^{11} m, and this is used as a convenient unit for the measurement of distances in the Solar System, called the **astronomical unit** (AU).

QUESTION 2.9

What are the average distances in AU from the Sun to (a) Mercury, and (b) Pluto? What are the average distances in units of $R_\odot$ from the Sun to (c) Mercury, and (d) the Earth?

2.4.1 The solar wind

In addition to episodic coronal mass ejections, there is also a continuous outflow of solar material that is termed the solar wind. The existence of a low-density, yet high-speed solar wind was proposed in 1951 by Ludwig Biermann as a result of a study of the observed shapes of the tails of comets. This suggestion was followed by an analysis of the structure of the solar corona by Eugene Parker. In an attempt to describe mathematically the structure of the corona, Parker suggested that the atmosphere around the Sun could not simply be static, and that material in the corona must flow out into interplanetary space.

The existence of the solar wind was confirmed directly by measuring its flow using experiments on Russian and American space probes in the mid-1960s. The temperature of the gas in the solar wind is typically very high, being over 10^5 K, which leads to the gas being highly ionized and hence forming a plasma. The composition of the solar wind is dominated by electrons and protons, along with a small fraction of helium nuclei and an even smaller percentage of heavier nuclei. The density of material in the solar wind is very low; in the vicinity of the Earth the number density of protons in the solar wind is about 7×10^6 m^{-3}. The speed of the wind is high, being typically several hundred km s^{-1}. One of the surprising results that emerged from the experiments that measured the solar wind directly, was that the speed of the wind seems to vary between a rather steady fast flow of about 750 km s^{-1}, and a much more erratic slower flow that has a speed of between 300 and 400 km s^{-1}.

As has already been mentioned, the origin of the solar wind in the solar corona was something that was anticipated before direct measurements of the wind itself were available. While the exact mechanism that drives the solar wind is still an area of active research, it seems clear that an important process involved in forming the solar wind arises from the relatively high gas pressure in the corona. Gas in the corona accelerates to reach outflow speeds of a few 100 km s^{-1} within a distance of about 20 to 30$R_\odot$ from the Sun, but then only accelerates slowly at distances beyond this (see Figure 2.33). This then is a partial explanation for the origin of the solar wind; the outflow from the corona simply carries on into interplanetary space. However, to understand why the speed of the solar wind (as measured near the Earth) seems to be either fast or slow, we need to consider the effect of the magnetic field of the Sun on gas in the corona.

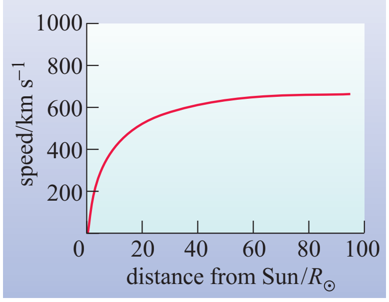


Figure 2.33 The speed of the solar wind away from the Sun increases as the distance from the Sun increases. The wind accelerates rapidly in the first few tens of $R_\odot$, and accelerates only slowly after this.

You saw in Section 2.3.5 that the magnetic field at the surface of the Sun varies considerably over the solar cycle. At times close to solar minimum the field over the surface of the Sun is weak but its pattern is similar to the dipole field. Close to solar maximum the field is much more complex and shows an irregular pattern that is difficult to characterize. We will concentrate here on how the solar magnetic field affects the solar wind around the minimum of the solar cycle.

- ■ What are the two types of interaction between a magnetic field and a plasma that you have already come across?
- ❑ In cases where the magnetic field is strong and the flow of plasma is weak, any movement of plasma is along field lines. In cases where the field is weak and the flow is strong, the field lines move with the plasma.

The structure of the magnetic field of the corona at solar minimum shows both these types of behaviour as illustrated in Figure 2.34. As mentioned above, at the surface of the Sun the field follows the dipole pattern closely. Those field lines that emanate from the *polar* regions pass high into the corona and into the region in which the bulk motion of the solar wind starts. These magnetic field lines are then carried by the solar wind and distort the dipole pattern, such that these field lines become highly extended as they are carried away from the Sun. The field lines that emerge from the Sun close to its magnetic *equator* form loops that do not reach high into the corona. These magnetic field lines may be distorted but they remain as closed loops, and ionized gas is trapped within these regions.

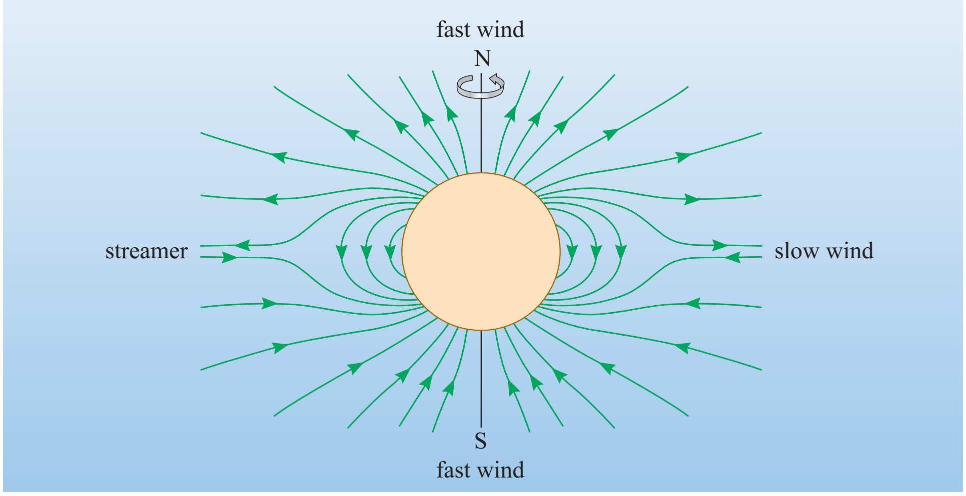


Figure 2.34 The magnetic field structure (green lines) of the solar corona and solar wind at a time close to the minimum of solar activity, showing the expected location of streamers and the location of the origin of fast and slow components to the solar wind. (Adapted from Pneuman and Kopp, 1971)

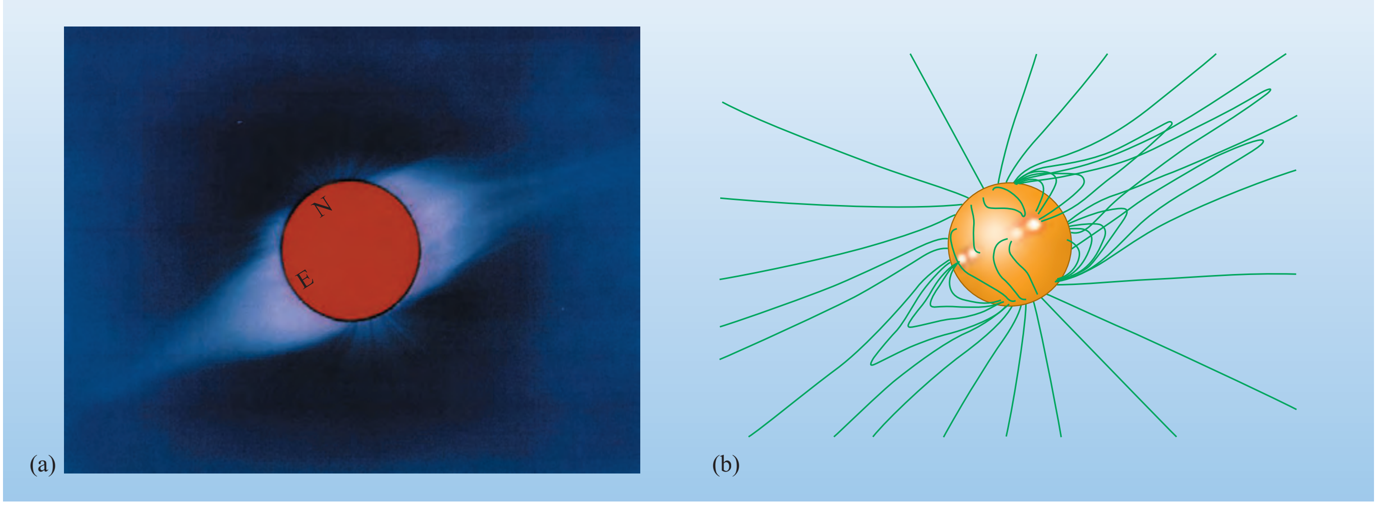


Figure 2.35 The solar corona during a total solar eclipse, at a time when the Sun is close to the minimum of solar activity: (a) in visible light; and (b) the magnetic field pattern that is deduced from the observed streamers. (The arrows showing the direction of the field lines have been omitted for clarity.) ((a) S. Koutchmy, Institut d'Astrophysique); (b) SAIC)

These structures provide an observational test for the model of how the solar magnetic field interacts with the corona and solar wind. Images of the corona taken during total solar eclipses near the minimum of solar activity often show 'helmet' streamers (so-called because their shape resembles old-fashioned pointed helmets), such as those shown in Figure 2.35a. The field pattern that is deduced from the shape of these streamers, as shown in Figure 2.35b, has similar features to the field pattern that is expected when a dipolar field interacts with an out-flowing plasma, in that some loops of field lines are distorted, but remain close to the Sun, whilst other loops become vastly extended as they are carried by the solar wind.

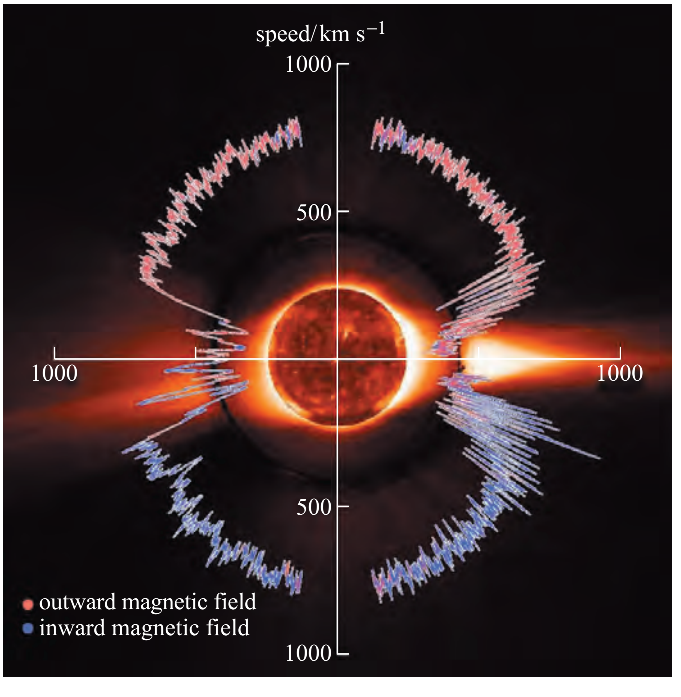


Figure 2.36 The solar wind speed as a function of heliocentric latitude as measured by the *Ulysses* spacecraft from 1992 to 1997. (Ulysses (ESA and NASA))

The fast component of the solar wind appears to originate in those parts of the corona where field lines do not loop back on themselves. These regions of open field lines correspond to the coronal holes that were mentioned in Section 1.4.3. Those regions of the Sun that are covered in closed magnetic loops seem to give rise to the more erratic slow component of the solar wind. This relationship between the two speeds of the solar wind and the magnetic field pattern was demonstrated using measurements made with the *Ulysses* space probe in the early 1990s. Most of the space probes that have explored the Solar System have stayed close to the equatorial plane of the Sun. *Ulysses* is remarkable in that its orbit was planned so that it would pass over the poles of the Sun, and therefore be used to measure the properties of the solar wind at different solar latitudes. The results of these measurements are displayed in Figure 2.36, which shows that at high latitudes, only the fast component of the solar wind is seen. However, near the solar magnetic equator, the solar wind shows the erratic behaviour that is characteristic of the slow component. Finally it should be noted that the magnetic axis and the rotation axis are not exactly aligned: there is an offset of several degrees between the two. This means that the solar wind which reaches the Earth (which orbits the Sun in a plane close to the solar equator) alternates between the fast and slow components.

QUESTION 2.10

Based on the density and the speed of the solar wind measured near the Earth, estimate the mass loss from the Sun due to the solar wind. Express your estimate in terms of $\mathrm{kg\,s^{-1}}$ and $M_\odot\,\mathrm{yr^{-1}}$. Assume that the density of the solar wind is 7×10^6 protons $\mathrm{m^{-3}}$ (ignore the contribution to the mass from other particles) and that the speed of the wind is that which corresponds to the 'fast' component. (*Hint*: calculate the mass of material that flows out over a sphere with radius equal to the radius of the Earth's orbit.)

2.4.2 Solar – terrestrial interactions

We have seen that the Sun is a source of out-flowing material, either in the form of the solar wind or coronal mass ejections, and it is of particular interest to us to understand the consequences of the interaction between this material and the Earth. The importance of the role of magnetic fields in determining the properties of the solar wind was illustrated in Section 2.4.1, and so it should not be surprising to find that the magnetic field of the Earth plays a key role in determining how the solar wind interacts with the Earth. Consider the situation that is likely to arise when a steady solar wind impinges on the magnetic field of the Earth. The fact that the solar wind is an ionized gas means that it cannot easily cross magnetic field lines. Therefore, the Earth's magnetic field presents an obstacle to the flow of the solar wind and, as in the solar corona, this results in either material being channelled along field lines, or the field lines being swept along with the wind. Upstream from the Earth there is a region where the pressure of the solar wind is similar to the resistance to the wind provided by the Earth's magnetic field. At this point the wind 'feels' the influence of the magnetic field and forms a feature called the **bow shock** (so-called because of its similarity to the bow wave that builds up in front of a moving ship). The flow of the wind is diverted around the Earth, and in doing so, the magnetic field pattern is swept back into a long tail as shown in Figure 2.37. The tear-drop shaped region around the Earth in which the Earth's magnetic field dictates how charged particles will move is called the **magnetosphere** (the name is somewhat misleading as it is not a spherical region). For most of the time, the

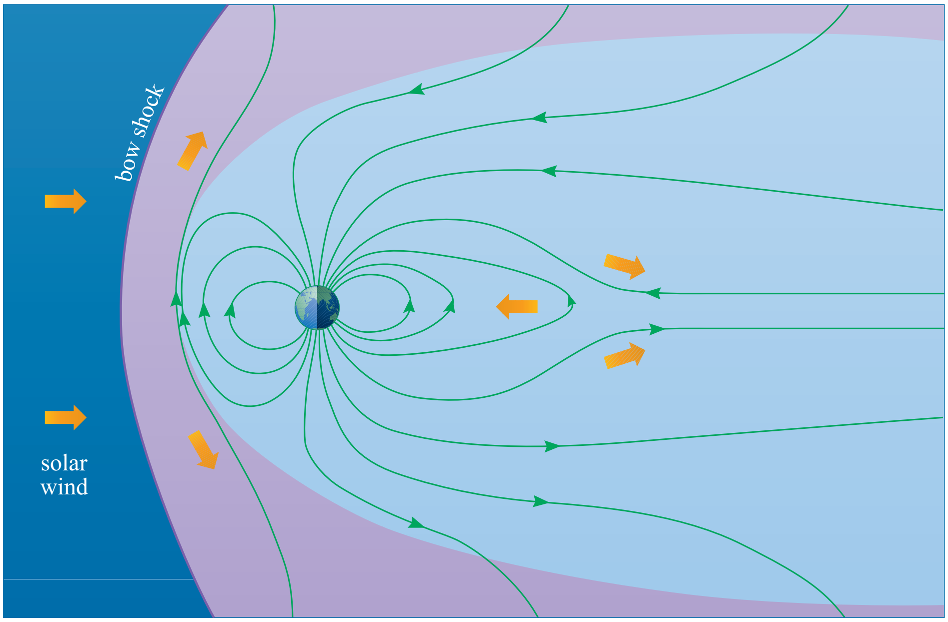


Figure 2.37 The magnetosphere of the Earth. The impact of the solar wind on the Earth's magnetic field distorts the field into the shape shown. The solar wind cannot easily cross magnetic field lines and so is 'swept' around the outer boundary of the magnetosphere (blue). However, some particles from the solar wind can enter through the 'tail' region of the magnetosphere – off to the right of the region shown in this diagram.

magnetosphere is a barrier that excludes most of the solar wind, although a small number of particles (electrons, protons and ionized atoms) continuously enter through the tail region. A further source of particles in the magnetosphere is the upper atmosphere of the Earth.

The extent of the magnetosphere depends on the balance between the force imparted by the solar wind and the resistance offered by the Earth's fixed magnetic field.

- ■ If the force imparted by the solar wind on the magnetosphere were to increase, what would happen to the position of the bow shock?
- ❑ The bow shock would move inwards towards the Earth, to a point where the force imparted by the wind is balanced by the stronger magnetic field closer to the Earth.

So the magnetosphere is a dynamic entity: its structure depends on the strength of the solar wind. Changes in the shape of the magnetosphere have an effect on the motion of charged particles that are trapped within it, and this in turn can lead to observable effects. One such effect is that a change in the shape of the magnetosphere can give rise to a variation in the local magnetic field at the surface of the Earth. Such variations are termed **geomagnetic disturbances** or, in extreme cases **geomagnetic storms**. Intense variations arise when large coronal mass ejections strike the magnetosphere. At times close to the maximum of the solar activity cycle, about ten such coronal mass ejections strike the Earth's magnetosphere every year and result in strong geomagnetic storms. The motion of charged particles in the magnetosphere can also give rise to very rapidly changing magnetic fields at the surface of the Earth, and this can induce currents in electrical power lines, leading to overloading and damage.

An important factor that determines the severity of a geomagnetic storm is the orientation of the magnetic field in the coronal mass ejection with respect to that of the Earth. We have already seen that CMEs are outflows of plasma, hence they transport magnetic field lines with them as they move away from the Sun.

- ■ Consider a coronal mass ejection that strikes the Earth's magnetosphere (i.e. moving to the right from the left-hand edge of Figure 2.37). If the magnetic field lines in the leading edge of the CME have opposite polarity to the Earth's magnetic field, what effect do you think may occur?
- ❑ Since magnetic field lines of opposite polarity are being forced together, it is possible that magnetic reconnection may occur (Section 2.3.4).

This is indeed what happens in the most severe geomagnetic storms – magnetic reconnection results in a rapid reconfiguration of the magnetic field, which causes the acceleration of electrons and protons that are located within the magnetosphere. The reconnection event also creates a breach in the magnetic barrier around the Earth, and allows particles from the CME to enter the magnetosphere.

Geomagnetic storms have several potentially adverse effects on human activity. Apart from the possibility of disruption of power supplies, the rapidly changing magnetic fields can also interfere with radio communication and navigation systems. Furthermore, the presence of energetic particles within the magnetosphere poses a health risk to astronauts and can also damage sensitive electronic equipment

on satellites. An increase in the number of energetic particles in the magnetosphere may arise from geomagnetic storms, but also from solar flares. Given the practical importance of knowing, or being able to predict the effect of solar activity on the Earth's environment, the state of solar–terrestrial interactions is continually monitored and reported under the generic name of **space weather**.

The most dramatic effect that can arise when the magnetosphere is disturbed is the production of an **aurora** (plural aurorae) which are also called the Northern or Southern Lights (aurora borealis and aurora australis, respectively). Aurorae are typically observed at locations on the Earth that have latitudes of 60 to 70 degrees (North or South), although they are occasionally seen from much lower latitudes. The form of an aurora usually resembles a glowing curtain of light (Figure 2.38) which seems to move and change in a matter of minutes. Aurorae themselves are observed to last for several hours, and in extreme cases for days.

Figure 2.38 An aurora. (Lionel Stevenson/Science Photo Library)

Aurorae are formed when the varying magnetic field of the magnetosphere causes electrons and protons from the tail region to move along the magnetic field lines down into the atmosphere of the Earth. These charged particles collide with atoms in the upper atmosphere, at heights of between about 100 km and 400 km above the surface of the Earth.

■ What may happen to an atmospheric atom as a result of these collisions?

❑ The atom may be excited to a higher electronic state, or if the collision is energetic enough the atom may become ionized.

Photons are emitted as these excited atoms or ions revert to lower electronic states, and it is these photons that form the visible auroral glow. Thus the spectrum of an aurora shows prominent emission lines from those elements that are common in the Earth's atmosphere, such as oxygen (green and red emission) and nitrogen (purple and red emission).

The particles that give rise to aurorae are channelled down magnetic field lines, and tend to impinge on the Earth's atmosphere in oval shapes that surround each magnetic pole. The auroral ovals (Figure 2.39) have been imaged in their entirety from satellites in orbits around the Earth, and show the huge extent of the aurora – a display seen from the ground is just a tiny part of the full oval.

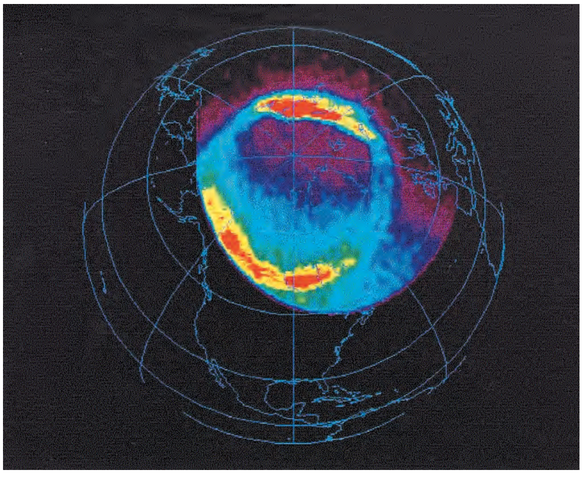


Figure 2.39 The northern auroral oval, as imaged by the POLAR satellite in ultraviolet emission. (G. E. Parks, University of Washington and the POLAR UVI team/NASA)

QUESTION 2.11

Discuss whether you might expect the Earth's aurorae to have prominent spectral lines due to the following: (a) atomic oxygen, (b) atomic iron, (c) ionized nitrogen.

2.4.3 The heliosphere

The solar wind streams outwards from the Sun into interplanetary, and eventually, interstellar space. As the solar wind gets further and further away from the Sun, its density drops while its speed remains roughly constant. The overall effect is that the pressure that the solar wind has due to its motion will drop with increasing distance from the Sun. So far, we have not discussed whether there is any gas in the space between the stars, but we can speculate that if there were, this would presumably have a pressure *of its own* due to the thermal motion of the particles in the gas. We will see later in this book that the space between the stars does indeed contain material: a mixture of gas and dust that is called the interstellar medium. Because the pressure of the solar wind continues to drop with distance from the Sun, at some point this pressure will balance the pressure of interstellar gas. The volume of space in which the pressure of the solar wind exceeds the pressure of gas in interstellar space defines the extent of the Sun's influence on its environment, a region that is called the **heliosphere**.

The boundary of the heliosphere is expected to have components as illustrated schematically in Figure 2.40. In moving outwards from the Sun, the solar wind flows freely until it reaches a region called the **termination shock** where it responds to the pressure of the interstellar medium by slowing down considerably. Further out from this shock is the boundary of the region in which the solar wind dominates the motion of matter. This boundary is termed the **heliopause**; beyond this boundary lies the interstellar medium. If the Sun is moving relative to the local

interstellar gas, then it is also likely that a bow shock will form ahead of the heliopause in the direction of motion, and that there would be a build up of interstellar gas in this region.

Observational data that could confirm this view and allow space scientists to determine the extent of the heliosphere are currently (2002) rather limited. It is believed that the Sun is moving relative to the local interstellar medium with a speed of 26 km s^{-1}, and hence it is anticipated that there is a bow shock and a density enhancement in the direction of the Sun's motion. The analysis of radio signals that are believed to originate when intense bursts of solar wind collide with interstellar gas, suggests that the heliopause may be between 110 and 160 AU from the Sun. The location of the termination shock is believed to be at about 85 AU from the Sun. It is hoped that the *Voyager 1* space probe, which had reached a distance of 83 AU from the Sun at the start of 2002, will pass through the termination shock within the next few years, and that within a few decades it will become the first artificial object to enter interstellar space.

In November 2003 contradictory findings about whether *Voyager 1* had passed the termination shock were published by two separate research collaborations. On the basis of one type of measurement, one group claimed that *Voyager 1* had passed the termination shock in mid-2002. The other group, who analysed different data, concluded that the spacecraft had yet to reach this feature of the heliosphere.

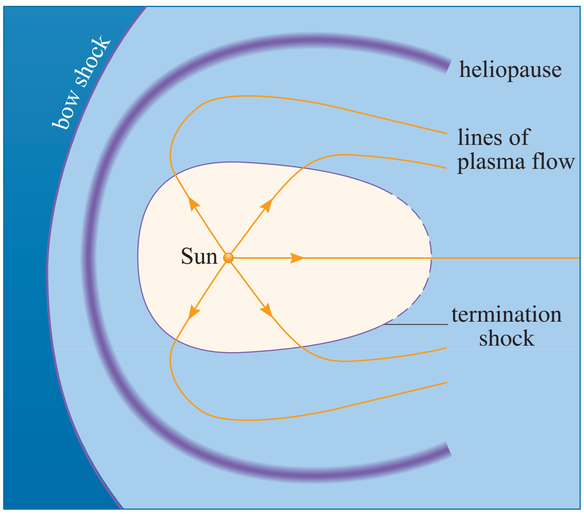


Figure 2.40 A schematic illustration of the expected structure of the heliosphere. (Adapted from Gosling, 1999)

QUESTION 2.12

Estimate the time taken for the fast component of the solar wind to travel from the Sun to the termination shock. Express your answer in days.

2.5 Summary of Chapter 2

The solar interior

- The temperature and density of the Sun increase as the distance from the centre (fractional radius) decreases. The core temperature is thought to be about 15.6×10^6 K.
- Theoretical models of the solar interior provide detailed structural information and favour a composition that is (by mass) roughly 73% hydrogen, 25% helium and 2% heavy elements, except in the central core, where the high temperatures promote chains of nuclear reactions that convert hydrogen nuclei into helium nuclei.

- The Sun's electromagnetic radiation originates in the central core, where it is released by the nuclear reactions that convert hydrogen nuclei into helium nuclei. In addition to high-energy γ-rays, these reactions also produce neutrinos.
- Energy released in the core makes its way to the surface by radiation and convection. Radiative energy transport in the solar interior is a gradual diffusive process in which the radiation is in local thermodynamic equilibrium with the material through which it passes.
- The solar neutrino problem is a deficit of the observed number of electron neutrinos in comparison to that expected from the solar model. The most likely explanation for this is that neutrinos change type as they travel from the core of the Sun to the Earth. This phenomenon is not well understood.
- The broad features of theories of the solar interior are not currently in dispute. However, observational tests involving solar oscillations and other sources of data may require some refinement in the details of those theories.

Solar activity

- The phenomena that indicate solar activity are primarily magnetic in origin. At times of high solar activity the Sun exhibits many active regions in which there are strong bipolar fields.
- Solar magnetic fields store energy that originates in the motion of the plasma in the surface layers in the Sun. This energy can be released rapidly giving rise to solar flares and coronal mass ejections. It is likely that magnetic reconnection plays a role in the process of energy release required for these type of events.
- The solar activity cycle is a cycle in which the global magnetic field of the Sun changes dramatically. The period taken for the magnetic field to return to its initial state is 22 years or twice the period inferred from sunspot number variations.

The solar wind and the heliosphere

- The corona of the Sun is not a static atmosphere, but undergoes a continuous outflow which forms the solar wind. The out-flowing material is a plasma, the composition of which is similar to the composition of the Sun, and the speed of flow, as measured near the Earth, is between about 300 and 750 $\mathrm{km\,s^{-1}}$.
- Out-flowing coronal plasma interacts with the magnetic field of the Sun. Depending on the strength of the magnetic field, this interaction may result in the flow being restricted to move along field lines or in the transport of field lines with the plasma.
- The magnetic field of the Earth acts as a barrier to exclude most of the solar wind from a region around the Earth. This region, called the magnetosphere, may shrink or expand as a result of changes in the strength of the solar wind, and this alters the pattern of flow of charged particles within the magnetosphere. This may give rise to geomagnetic storms and aurorae.
- The pressure exerted by the solar wind drops with distance from the Sun, and eventually balances the pressure of interstellar gas. This boundary marks the limit of the heliosphere, and is thought to lie between 110 and 160 AU from the Sun.

Questions

QUESTION 2.13

Estimate the numbers of hydrogen and helium nuclei contained in the Sun.

QUESTION 2.14

Estimate the total amount of time for which the solar luminosity can be sustained at its current value by the conversion of hydrogen into helium.

QUESTION 2.15

Why are the following nuclear reactions impossible?

(a) ${}^{1}_{1}\mathrm{H} + {}^{4}_{2}\mathrm{He} \rightarrow {}^{5}_{3}\mathrm{He}$

(b) ${}^{7}_{4}\mathrm{Be} + {}^{1}_{1}\mathrm{H} \rightarrow 2\,{}^{4}_{2}\mathrm{He} + 2\mathrm{e}^{+} + \gamma + 2\nu_{\mathrm{e}}$

(c) ${}^{3}_{2}\mathrm{He} + {}^{4}_{2}\mathrm{He} \rightarrow {}^{7}_{3}\mathrm{Li} + \gamma$

QUESTION 2.16

If the Sun's luminosity is entirely due to the loss of rest energy from the solar core, how much mass does the Sun lose each year?

QUESTION 2.17

Calculate the number of photospheric photons that are required to carry away the energy generated in a single ppI reaction.

QUESTION 2.18

The structure of filaments is an area of on-going investigation. One model that has been proposed to explain how a filament is supported is based on the idea that a cross-section of the filament (perpendicular to the direction in which it is elongated) may have a field pattern as shown in Figure 2.41. Explain how this field pattern supports the filament.

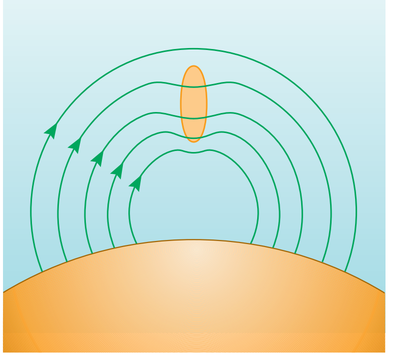


Figure 2.41 A proposed model for the way in which a filament may be supported by a magnetic field. The field pattern is shown for a plane that is perpendicular to the direction of elongation of the filament. For use with Question 2.18. (Phillips, 1992)

QUESTION 2.19

Briefly describe the energy conversions that may occur in magnetic reconnection.